

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO SOCIAL

SILVIA DE FREITAS DAL BEN FURTADO

**CARTOGRAFANDO O JORNALISMO AUTOMATIZADO:
REDES SOCIOTÉCNICAS E INCERTEZAS NA REDAÇÃO
DE NOTÍCIAS POR “ROBÔS”**

BELO HORIZONTE

2018

SILVIA DE FREITAS DAL BEN FURTADO

**CARTOGRAFANDO O JORNALISMO AUTOMATIZADO:
REDES SOCIOTÉCNICAS E INCERTEZAS NA REDAÇÃO
DE NOTÍCIAS POR “ROBÔS”**

Dissertação apresentada ao Programa de Pós-Graduação em Comunicação Social da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Comunicação Social.

Linha de Pesquisa: Textualidades Midiáticas

Orientador: Prof. Dr. Carlos Frederico de Brito d'Andréa

BELO HORIZONTE

2018

301.16 Dalben, Silvia de Freitas
D137c Cartografando o jornalismo automatizado [manuscrito] :
2018 redes sociotécnicas e incertezas na redação de notícias por
"robôs" / Silvia de Freitas Dal Ben Furtado. - 2018.
117 f.
Orientador: Carlos Frederico de Brito D'Andréa.

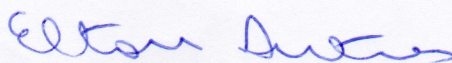
Dissertação (mestrado) - Universidade Federal de Minas
Gerais, Faculdade de Filosofia e Ciências Humanas.
Inclui bibliografia.

1.Comunicação – Teses. 2. Comunicação de massa –
Teses. 3.Jornalismo - Teses 4.Repórteres e reportagens –
Teses. 5.Robôs - Teses. I. D'Andréa, Carlos Frederico de
Brito. II. Universidade Federal de Minas Gerais. Faculdade
de Filosofia e Ciências Humanas. III. Título.

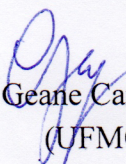
*Cartografando o jornalismo automatizado: redes sociotécnicas e incertezas na
redação de notícias por “robôs”
(trabalho orientado pelo Prof. Dr. Carlos Frederico de Brito D’Andrea)*

Sílvia de Freitas Dal Ben Furtado

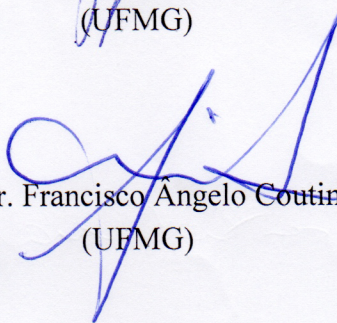
Dissertação defendida e aprovada pela banca examinadora:



Prof. Dr. Elton Antunes
(UFMG)



Profa Dra. Geane Carvalho Alzamora
(UFMG)



Prof. Dr. Francisco Angelo Coutinho
(UFMG)

Programa de Pós-graduação em Comunicação Social
Faculdade de Filosofia e Ciências Humanas
Universidade Federal de Minas Gerais
Belo Horizonte, 20 de junho de 2018.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO SOCIAL
Av. Presidente Antonio Carlos, 6627
Prédio da Fafich, sala 4232 – 31270.901
Fone/Fax: 31 34095072 E- mails: ppgcom@fafich.ufmg.br;
Home-page: www.fafich.ufmg.br/ppgcom



DECLARAÇÃO

Declaramos que *Sílvia de Freitas Dal Ben Furtado*, em 20 de junho de 2018 (1º semestre de 2018), defendeu seu trabalho de dissertação, intitulado **Cartografando o jornalismo automatizado: redes sociotécnicas e incertezas na redação de notícias por “robôs”** junto ao Programa de Pós-graduação em Comunicação Social da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais. O trabalho de dissertação foi orientado pelo Prof. Dr. Carlos Frederico de Brito D’Andrea e aprovado pelos membros da banca examinadora composta pelos professores doutores Elton Antunes (Universidade Federal de Minas Gerais), Geane Carvalho Alzamora (Universidade Federal de Minas Gerais), e Francisco Ângelo Coutinho (Universidade Federal de Minas Gerais).

Belo Horizonte, 20 de junho de 2018.

PROF. DR. CARLOS MAGNO CAMARGOS MENDONÇA

Coordenador do Programa de Pós-Graduação em
Comunicação Social - FAFICH/UFMG

AGRADECIMENTOS

Eu gostaria de agradecer a meus pais por não terem medido esforços durante toda a minha vida para investirem em minha educação, e por estarem sempre ao meu lado. Ao Arthur, pelo apoio incondicional e por não me deixar desistir nunca de meus sonhos. Ao Guilherme, por compreender a ausência da mãe em vários momentos e sempre falar com carinho dos nossos amigos “robôs jornalistas”.

À minha irmã Mirian pelo exemplo de dedicação, excelência e profissionalismo que sempre me guiaram. À minha avó Tininha, pela força e inspiração de manter sempre o sorriso no rosto, apesar das adversidades que a vida nos impõe. À minha família, em especial à minhas tias e primas, pelo exemplo e pelo carinho. Sou privilegiada de ter nascido no meio de tantas mulheres fortes e vencedoras.

Ao meu orientador Prof. Carlos d’Andréa, por ter embarcado nessa aventura comigo, e ter se dedicado com tanto afinho a minha pesquisa. Muito obrigada pelos ensinamentos, tanto nas atividades de pesquisa quanto de docência, e por estar sempre presente apesar da distância física destes últimos meses.

Às amigas Polyana Inácio, Amanda Jurno e Verônica Soares, que me ajudaram inúmeras vezes a superar as dificuldades que surgiram ao longo do percurso acadêmico. Vocês são fontes inesgotáveis de inspiração.

À Universidade Federal de Minas Gerais, ao Programa de Pós-Graduação em Comunicação Social e a todos os professores que me ajudaram nesta trajetória, em especial ao Prof. Carlos Falci que foi o primeiro a me motivar a me dedicar a este tema em 2013, em uma disciplina isolada na Escola de Belas Artes. Agradeço ainda aos professores Francisco Coutinho, Elton Antunes, Joana Ziller, Geane Alzamora, Fernanda Duarte, Carlos Mendonça, Lorena Tárzia e Ana Carolina Vimieiro, cujos comentários e observações me ajudaram a refletir sobre o meu objeto de pesquisa. Gostaria de agradecer também aos professores Adriano Veloso e Nívio Ziviane, que em 2015 elucidaram várias dúvidas técnicas sobre este objeto e sobre *Natural Language Processing*.

À Fundação de Amparo à Pesquisa de Minas Gerais pela bolsa de estudos que me permitiu dedicar os últimos dois anos a esta pesquisa.

A todos os colegas do NucCon pelas boas discussões das quais participei nestes últimos dois anos, em especial a Luciana Andrade, Leonardo Melgaço, Tiago Salgado, André Mintz, Pedro Nogueira, Bruno Fonseca, Vitória Barros e Emmanuele Araújo.

Ao amigo Gary Krieger, por todos os conselhos, por todas as conversas sobre programação em R e em *Python*, por me inspirar a ampliar as perspectivas tanto acadêmicas quanto profissionais.

Aos jornalistas e profissionais de T.I com quem trabalhei no Portal Uai que, de maneira indireta, influenciaram a trajetória desta pesquisa: Benny Cohen, Rodrigo Rocha, Carolina Braga, João Renato Faria, Roque Viana e Fred Bottrel.

Por fim, não poderia deixar de me lembrar dos vários não-humanos que me ajudaram nesta caminhada cercada de híbridos. Não conseguirei citar todos, mas gostaria de agradecer ao meu computador *Apple* e ao *Microsoft Word*, que foram companheiros diários neste trabalho; ao *Google* e suas várias ferramentas de busca, com destaque para o *Google Acadêmico*, *Google Tradutor* e *Google Patentes*, que me permitiram ter acesso a informações produzidas em terras distantes e mapear alguns dos atores envolvidos no Jornalismo Automatizado. Vocês representam um universo enorme de atores não-humanos e dos inúmeros algoritmos que estiveram ao meu lado durante esta pesquisa.

Muito obrigada! Sem vocês, não teria conseguido desenvolver este trabalho.

“Intelligence is the ability to adapt to change.”

Stephen Hawking

RESUMO

Desde 2010, jornais de diversos países adotaram *softwares* de *Natural Language Generation* (NLG), um subcampo da Inteligência Artificial, para produzir notícias automatizadas sobre finanças, esportes, eleições, crimes e terremotos. Denominado como “robô repórter” (CARLSON, 2014), “notícias escritas por máquina” (VAN DALEN, 2012), “Jornalismo Automatizado” (CARLSON, 2014; GRAEFE, 2016) ou “Jornalismo Algorítmico” (DÖRR, 2015), este fenômeno é caracterizado pela publicação de milhares de notícias dirigidas por dados, com estrutura simples e repetitiva, criadas a partir de um *template* e um banco de dados estruturado acionados por algoritmos. Nesta pesquisa, buscamos inspiração no STS (*Science and Technology Studies* em inglês) e na Teoria Ator-Rede (TAR) para traçar as redes sociotécnicas mobilizadas pelo Jornalismo Automatizado, em um esforço de abrir as caixas-pretas e revelar os agenciamentos de atores humanos e não-humanos vinculados a esta inovação tecnológica. Incomodados com algumas visões que acreditam que as notícias automatizadas são produzidas sem intervenção humana, os conceitos de autonomia tecnológica (WINNER, 1977) e mediação técnica (LATOUR, 1994^b) nos auxiliam neste percurso que busca reconhecer que por trás destes textos existe uma complexa rede de jornalistas, programadores, analistas de dados, empreendedores, pesquisadores, *softwares*, algoritmos e bancos de dados. Para descrever os atores, a formação de redes e as associações vinculadas ao Jornalismo Automatizado, somos influenciados pelas lentes de observação sugeridas pela Cartografia de Controvérsias (VENTURINI, 2010, 2012) como um modo de olhar para esta inovação, o que fundamenta a nossa metodologia. Buscamos estabelecer um ponto de vista simétrico e não-antropocêntrico, que não recaia nem em uma perspectiva determinista, nem em uma visão pessimista, e tente compreender as incertezas desencadeadas por uma “suposta” autonomia tecnológica de “robôs” redatores de notícias. A cartografia desta pesquisa é dividida em dois movimentos. Em um primeiro movimento, buscamos situar o Jornalismo Automatizado em um contexto acadêmico, onde sugerimos que este seja um campo de estudos em formação que dialoga com as Ciências da Computação e com o Jornalismo Computacional. Nesta etapa da cartografia, descrevemos como funcionam os *softwares* NLG e alguns experimentos que analisam como jornalistas e leitores reagem a esta tecnologia. Em um segundo movimento, apresentamos três estudos de caso de veículos jornalísticos que adotaram algoritmos para a produção automatizada de notícias. O primeiro relata a trajetória da *startup Narrative Science*, incluindo as parcerias com a *Forbes* e a *ProPublica* e a controversa afirmação de que em breve um computador ganharia o *Pulitzer*. O segundo descreve o percurso da *startup Automated Insights*, com o lançamento da rede *StatSheet* e a parceria com a agência *Associated Press*. O terceiro apresenta dois projetos desenvolvidos pelo *Los Angeles Times*: o *The Homicide Report* para a cobertura de homicídios e *Quakebot* que agiliza a publicação de notícias sobre terremotos. A partir desta cartografia, argumentamos que existe um novo ecossistema em formação onde jornalistas cada vez mais trabalham ao lado de engenheiros, programadores, analistas de dados. Nesta pesquisa, defendemos que os *softwares* NLG não eliminam, porém desafiam os jornalistas a se adaptarem para lidar com os dados de uma nova forma, em um contexto onde os algoritmos, o *big data* e as linguagens de programação se tornam cada vez mais presentes.

Palavras-chave: Jornalismo Automatizado, robô repórter, *Science Technology Studies*, Teoria Ator-Rede, *Natural Language Generation*

ABSTRACT

Since 2010, newspapers from several countries adopted Natural Language Generation (NLG) softwares, a subfield of Artificial Intelligence, to produce automated news about finance, sports, elections, crimes, and earthquakes. Named as "the robot reporter" (CARLSON, 2014), "machine-written news" (VAN DALEN, 2012), "Automated Journalism" (CARLSON, 2014, GRAEFE, 2016) or "Algorithmic Journalism" (DÖRR, 2015), this phenomenon is characterized by the publication of thousands of data-driven news stories, with simple and repetitive structure, created with a template and a structured database guided by algorithms. In this research, we sought inspiration from STS (Science and Technology Studies) and the Actor-Network Theory (TAR) to map the sociotechnical networks mobilized by Automated Journalism in an effort to open the black boxes and reveal the assemblages of human and non-human actors linked to this technological innovation. Refusing some views which believe automated news are produced without human intervention, the concepts of autonomous technology (WINNER, 1977) and technical mediation (LATOUR, 1994) help us in this process, which seeks to recognize that automated news are produced by a complex network of journalists, entrepreneurs, programmers, data scientists, softwares, algorithms and databases. To describe the actors, the networks arrangements and the associations linked to Automated Journalism, we are influenced by the observation lenses suggested in the Cartography of Controversies (VENTURINI, 2010, 2012) as a way of looking at this innovation, which underlies our methodology. We seek to establish a symmetrical and non-anthropocentric point of view that does not fall into either a deterministic or pessimistic perspective, and attempts to understand the uncertainties triggered by a "supposed" technological autonomy of news written by "robots". The cartography of this research is divided into two movements. At first, we seek to situate Automated Journalism in an academic context, where we suggest that this is a field of study in formation that dialogues with Computer Science and Computational Journalism. In this stage of the cartography, we describe how NLG software works and some experiments that analyses how journalists and readers react to this technology. In a second movement, we present three case studies about journalistic vehicles that adopted algorithms to automate news production. The first recounts the trajectory of the startup Narrative Science, including its partnerships with Forbes and ProPublica and the controversial statement that a computer would soon win the Pulitzer Prize. The second describes the path of the startup Automated Insights, from the launch of StatSheet network to its partnership with Associated Press. The third presents two projects developed by Los Angeles Times: The Homicide Report focused on homicide coverage and Quakebot that accelerate earthquake news publication. From this cartography, we argue that a new ecosystem is being created where journalists are increasingly working with engineers, programmers, data analysts. In this research, we claim that NLG software does not eliminate, but challenges journalists to adapt and deal with data in a new way, in a context where algorithms, big data and programming languages become more and more present.

Keywords: Automated Journalism, robot reporter, Science Technology Studies, Actor-Network Theory, Natural Language Generation

LISTA DE FIGURAS

- Figura 1** Manchete de reportagens sobre notícias automatizadas
- Figura 2** Robôs humanoides ilustram reportagens sobre Jornalismo Automatizado
- Figura 3** Exemplo das primeiras previsões do tempo escritas por um computador
- Figura 4** Fluxogramas retirados de dois registros de patentes da *Narrative Science*
- Figura 5** Tabela CSV do *Wordsmith* com resultados financeiros de fundos de investimento
- Figura 6** Exemplo de *template* do *Wordsmith* para gerar relatórios sobre resultados financeiros de fundos de investimento a partir dos dados da tabela da Figura 5
- Figura 7** Percepção dos leitores para notícias escritas por *software* e por jornalista
- Figura 8** Leitores avaliam a expectativa e a percepção de notícias automatizadas em uma análise comparada com notícias escritas por um jornalista
- Figura 9** Exemplo de notícia automatizada escrita pelo *StatsMonkey*
- Figura 10** Notícia automatizada publicada na *Forbes*
- Figura 11** Notícia automatizada publicada pela *ProPublica* com o perfil de uma das 52 mil escolas que fazem parte do projeto *The Opportunity Gap*
- Figura 12** Arquivos em formato *JSON* das notícias automatizadas produzidas pela *Narrative Science* para o projeto *The Opportunity Gap* da *ProPublica*
- Figura 13** Email da *Narrative Science* confirma a mudança de rumo da empresa
- Figura 14** Notícia automatizada produzida pela *StatSheet*
- Figura 15** Notícia automatizada publicada pela *Associated Press*
- Figura 16** Gráfico compara notícias sobre o mercado financeiro publicadas pela *Associated Press* escritas por repórteres e por *software* entre 2012 e 2015
- Figura 17** Capa do *Homicide Report* no site do *Los Angeles Times*
- Figura 18** *Post* automatizado publicado no *Homicide Report*
- Figura 19** Notícia atualizada por jornalista no *Homicide Report*
- Figura 20** Email enviado pelo USGS com os dados do terremoto
- Figura 21** Notícia sobre o terremoto escrita pelo *Quakebot*
- Figura 22** Notícia sobre o terremoto após atualizações feitas por jornalista
- Figura 23** *Tweets* do *L.A. Times* se retratando por um erro do *Quakebot*

SUMÁRIO

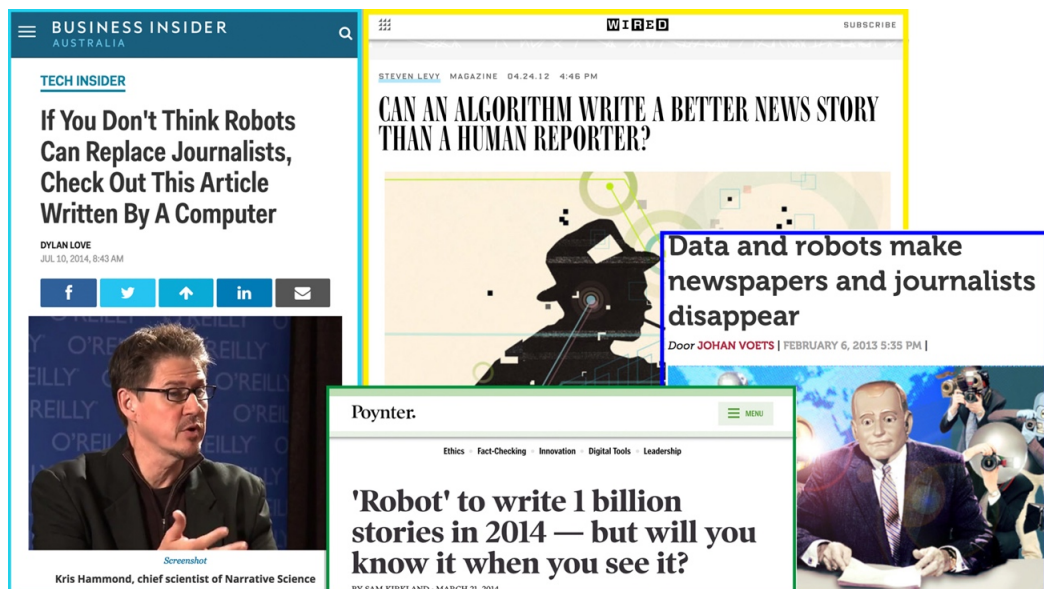
1. Introdução	10
1.1 Sobre a estrutura desta dissertação	16
2. Jornalismo pelo STS e pela TAR: interfaces e metodologias	18
2.1 <i>Science and Technology Studies</i> (STS) e Teoria Ator-Rede (TAR)	18
2.2 O jornalismo a partir da perspectiva do STS e da Teoria Ator-Rede	23
2.3 Duas concepções de autonomia: Winner e Latour	28
2.4 Cartografando as incertezas do Jornalismo Automatizado	34
2.5 Como cartografamos o Jornalismo Automatizado nesta pesquisa	39
3. Jornalismo Automatizado: cartografando tecnologias e pesquisas	42
3.1 Décadas de automação no jornalismo	42
3.2 Como funcionam os <i>softwares</i> de <i>Natural Language Generation</i>	45
3.3 <i>Softwares</i> NLG no jornalismo: experimentos e legitimação	50
3.4 Jornalismo Automatizado: um campo de estudos em formação	55
4. Jornalismo Automatizado: cartografando três estudos de caso	62
4.1 <i>Narrative Science</i> : Do <i>Pulitzer</i> à desistência do jornalismo	62
4.1.1 <i>StatsMonkey</i>	62
4.1.2 <i>Narrative Science</i>	66
4.1.3 Uma parceria com a <i>Forbes</i>	69
4.1.4 Projeto educacional com a <i>ProPublica</i>	72
4.1.5 Mudança de rumo e abandono do jornalismo	76
4.2 <i>Automated Insights</i> : <i>StatSheet</i> e a parceria com a <i>Associated Press</i>	77
4.2.1 <i>StatSheet</i>	77
4.2.2 Uma parceria com a <i>Associated Press</i>	79
4.3 Do jornalismo à automatização: a experiência do <i>Los Angeles Times</i>	84
4.3.1 Uma nova forma de reportar homicídios	84
4.3.2 <i>Quakebot</i> , o robô que escreve (mas não publica) sobre terremotos	89
4.4 O que os estudos de caso revelam sobre o Jornalismo Automatizado	94
5. Considerações Finais	102
Referências Bibliográficas	107

1 Introdução

Em março de 2017, a revista *ProXXI* do grupo *Meio & Mensagem* publicou uma matéria com um título alarmante: “Mande seus jornalistas embora, que meu software resolve o resto para você”. (MARCONDES, 2017) O texto repercutia duas palestras apresentadas no festival SXSW, em Austin nos Estados Unidos, que discutiam o futuro do Jornalismo Automatizado e projetos de automação desenvolvidos pelos jornais *The Washington Post* e *The New York Times*. O que chama atenção neste texto é o tom enfático e pessimista que descreve esta tecnologia como uma máquina que “cospe uma notícia pronta e acabou”, uma visão simplificada que propaga a ideia de que, para ter acesso a esta inovação, basta adquirir o software *Wordsmith* com um custo mensal de US\$ 1.000,00. “Imagine isso comparado com os salários de todos nós, jornalistas. É o começo do fim para todos nós.” (Idem)

Como esta, são várias as reportagens que adotam títulos sensacionalistas para divulgar a implementação de *softwares* que redigem notícias (Figura 1), publicadas principalmente a partir de 2012, quando a *Forbes* anunciou a adoção desta tecnologia para a redação de balancetes financeiros de empresas listadas na bolsa.

Figura 1 – Manchete de reportagens sobre notícias automatizadas



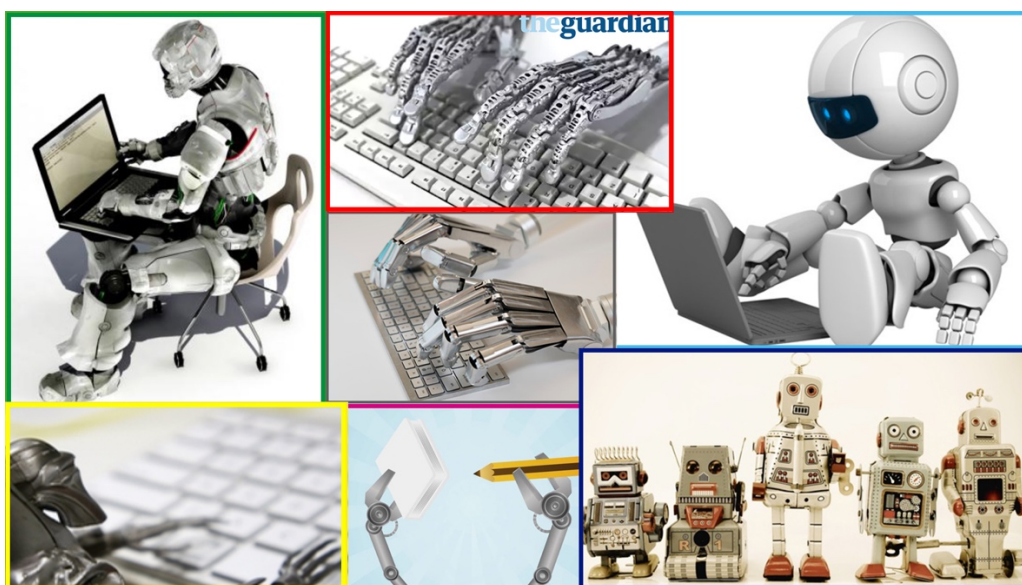
Fonte: Reprodução de internet¹

¹ Montagem criada pela pesquisadora a partir de imagens reproduzidas da internet. Fonte: <https://www.businessinsider.com.au/narrative-science-quill-gamechanger-2014-7> (LOVE, 2014); <https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/> (LEVY, 2012); <http://numrush.com/2013/02/06/data-and-robots-make-newspapers-and-journalists-disappear/> (VOETS, 2013); <https://www.poynter.org/news/robot-write-1-billion-stories-2014-will-you-know-it-when-you-see-it> (KIRKLAND, 2013).

Como mostra a Figura 1, as manchetes questionam se seria possível um algoritmo escrever melhor uma notícia que um repórter, exaltam a capacidade de um “robô” que escreveu um bilhão de notícias em 2014, e ainda levantam a hipótese de que “robôs” e dados serão os responsáveis pelo desaparecimento dos jornais impressos e dos jornalistas.

Extrapolando os títulos, um outro padrão narrativo que se destaca nestas reportagens é o recorrente uso de imagens com mãos robóticas sobre o teclado e robôs humanoides sentados na frente de um computador, utilizadas inúmeras vezes para ilustrar estes textos. (Figura 2)

Figura 2 – Robôs humanoides ilustram reportagens sobre Jornalismo Automatizado



Fonte: Reprodução de internet²

Ao associar o Jornalismo Automatizado a imagens de robôs humanoides, estas reportagens evocam um imaginário social pessimista que considera o avanço tecnológico uma ameaça para a humanidade, retórica frequentemente retratada em filmes e romances de ficção científica. Esta estratégia atrai a atenção do leitor, mas por outro lado invisibiliza várias especificidades do Jornalismo Automatizado, reiterando um discurso de medo e evocando

² Montagem criada pela pesquisadora a partir de imagens reproduzidas da internet. Fonte: http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html (OREMUS, 2014); <https://www.theguardian.com/media/shortcuts/2014/mar/16/could-robots-be-journalist-of-future> (GANI, HADDOU, 2014); <https://www.theatlantic.com/technology/archive/2012/03/robot-journalism-isnt-scary-its-just-plain-bad/330324/> (GREENFIELD, 2012); <http://www.pulzo.com/tecnologia/asi-funcionan-los-robots-periodistas-que-transforman-datos-en-textos/PP69774> (PULZO, 2016); <http://gizmodo.uol.com.br/por-que-devemos-nos-empolgar-com-a-invencao-de-robos-jornalistas/> (BURGOS, 2012); http://mashable.com/2014/07/01/robot-reporters-add-data-to-the-five-ws/#r4J2zh_3Oggh (ULANOFF, 2014); <https://www.theatlantic.com/technology/archive/2012/04/robot-journalism-still-doesnt-sound-scary/328893/> (GREENFIELD, 2012).

reações inflamadas de jornalistas que saem em defesa da profissão. Por exemplo, quando Marcondes (2017) sugere que esta tecnologia “*cospe uma notícia pronta e acabou*”, ele levanta a possibilidade de substituição de jornalistas por um *software*, no caso o *Wordsmith*. Porém, quando observamos o funcionamento deste sistema e os processos de produção de uma notícia automatizada, vemos uma realidade bem distante da ideia de robôs humanoides e mais próxima de computadores, *softwares* e algoritmos, objetos que já fazem parte do dia a dia das redações há algumas décadas.

Wordsmith é o nome dado ao *software* de *Natural Language Generation*³ desenvolvido pela *startup Automated Insights*, que realiza o processamento de grandes volumes de dados estruturados em tabelas CSV para gerar como *output* textos – que podem ser notícias, relatórios, documentos – que seguem uma estrutura repetitiva e pré-determinada. De forma simplificada, podemos dizer que *softwares* NLG funcionam como um tradutor, convertendo dados em textos. Desde 2014, o *Wordsmith* é utilizado pela agência *Associated Press* para a produção de cerca de cinco mil notícias sobre resultados trimestrais de empresas listadas na bolsa, uma experiência semelhante à adotada pela *Forbes* em parceria com outra *startup*, a *Narrative Science*.

Softwares como este estão sendo adotados por vários veículos ao redor do mundo desde 2010. O jornal *Los Angeles Times*, por exemplo, desenvolveu internamente com sua equipe de jornalismo de dados dois projetos que utilizam esta tecnologia: o *The Homicide Report* que publica uma notícia para cada homicídio reportado em Los Angeles, e o *Quakebot* que agiliza a publicação de notícias com informações iniciais sobre terremotos. Adquirido pela *Amazon* em 2013, o *The Washington Post* é outro veículo que também utiliza *softwares* para a produção de notícias automatizadas. Seu *software* denominado *Heliograf* foi desenvolvido internamente por uma equipe de engenheiros e utilizado, a partir de 2016, para a cobertura das Olimpíadas, das eleições presidenciais e de jogos de futebol americano.

Na Europa, a experiência de maior destaque foi realizada pelo jornal francês *Le Monde*, que adotou o *software Data2Content* desenvolvido pela *startup Syllabs* para a cobertura das eleições departamentais em 2015, quando foram publicadas 36 mil notícias automatizadas em poucas horas com os resultados do pleito de todos os municípios e cantões. Na Alemanha, existem também pelo menos quatro *startups* dedicadas ao desenvolvimento desta tecnologia, onde a que mais se destaca é a *Aexea*, desenvolvedora do *software AX Semantics* que promete redigir textos automatizados em 12 línguas. Há ainda

³ *Natural Language Generation* é um subcampo da Inteligência Artificial que abrange o desenvolvimento de sistemas com o intuito de produzir textos em linguagem natural. (REITER, DALE, 2000)

experiências na Escócia com a empresa *Arria NLG*, na Suécia com a *startup United Robots* adotada pelo grupo *MittMedia*, e na Rússia com a empresa *Yandex*.

Na Ásia, o Jornalismo Automatizado também é adotado na China em pelo menos quatro experiências: a *Tencent* desenvolveu o *software Dreamwriter*; enquanto a agência pública de notícias *Xinhua* publica notícias automatizadas de finanças e esportes produzidas pelo *software Kuaibi Xiaoxi*; o jornal *Southern Metropolis Daily* utiliza o *Xian Nan Robot* desenvolvido pelo professor universitário Wan Ziaojun, da *Peking University*; e o *China Earthquake Network Center* publica notícias automatizadas sobre terremotos. Na Coreia do Sul, a agência de notícias *Yonhap* utiliza o *software Soccerbot* desde 2017 para a cobertura de jogos de futebol da *English Primer League* (EPL).

No Brasil, em 2013 uma pesquisa acadêmica realizada na Universidade Federal do Maranhão pelo professor Márcio Carneiro dos Santos desenvolveu um *software* para a escrita de *leads* sobre resultados de partidas do Campeonato Brasileiro de Futebol (SANTOS, 2016). Desde 2017, um outro projeto que foi batizado de *Serenata do Amor* utiliza Inteligência Artificial para monitorar os gastos dos Deputados Federais que são reembolsados pela Cota para Exercício da Atividade Parlamentar (CEAP). Este projeto se intitula como “primeiro robô-jornalista do Brasil” e prevê para o segundo semestre de 2018, antes das eleições de outubro, começar a publicar pequenas notícias automatizadas a partir de informações sobre leis, patrimônio dos políticos, doações recebidas pelas campanhas, projetos de lei e gastos com cotas e emendas parlamentares. (MONNERAT, 2018)

Neste cenário, se de um lado vemos a expansão do Jornalismo Automatizado pelo mundo em iniciativas que adotam *softwares* NLG para coberturas de esportes, finanças, terremotos, eleições, homicídios, previsões do tempo, trânsito, de outro várias dúvidas surgem quanto ao uso desta tecnologia. Por exemplo, qual a dimensão e os limites técnicos para a aplicação destes *softwares* na rotina das redações jornalísticas? Quais são os impactos sociais, políticos e éticos que a implementação desta tecnologia gera? Qual a autoria e a responsabilidade legal de uma notícia automatizada? Será que esta tecnologia tem o potencial de substituir repórteres nas redações? Quais atividades os *softwares* NLG estão aptos a desempenhar? Como os repórteres reagem à introdução desta tecnologia e como eles conseguem diferenciar o seu trabalho daquele executado por um “robô”? Como os leitores avaliam as notícias automatizadas? Eles conseguem distinguir se um texto foi escrito por um repórter ou por um algoritmo?

A polêmica em torno da aplicação desta tecnologia no jornalismo é tão grande que muitos veículos que a adotam, por exemplo, preferem não revelar a seus leitores que determinados conteúdos foram escritos por um *software*, pois temem que isto chame a atenção e estimule uma reação mais crítica quanto aos textos. Um outro ponto levantado seria

um tratamento diferenciado pelos algoritmos de *ranking* em sistemas de busca como o *Google*, que poderiam qualificar de forma inferior as notícias automatizadas. (KIRKLAND, 2014)

Um dos fatores que contribuiu para aquecer este debate foi a declaração em 2011 de um dos fundadores da *Narrative Science* que afirmou que em cinco anos um programa de computador ganharia o *Pulitzer* (LOHR, 2011) – um aclamado prêmio norte-americano oferecido pela *Columbia University* anualmente para trabalhos jornalísticos de excelência – e que em 15 anos 90% das notícias seriam escritas de forma automatizada. (LEVY, 2012) Declarações como esta colaboram para difundir a ideia de que estas notícias são redigidas apenas por máquinas, sem nenhuma intervenção humana, o que pode colocar o emprego dos jornalistas em risco no caso destes *softwares* serem adotados de forma massiva pelos veículos de comunicação.

E esta visão também ganha repercussão na academia. Por ser um fenômeno recente, ainda não há um consenso em como denominar o uso desta tecnologia pelo jornalismo. Nesta pesquisa, optamos por utilizar os termos Jornalismo Automatizado (CARLSON, 2014; GRAEFE, 2016) e notícias automatizadas (CARREIRA, 2017), mas outros pesquisadores usam diferentes nomes como Dörr (2015) que adota a expressão “*Algorithmic Journalism*” (Jornalismo Algorítmico), ou Van Dalen (2012) que prefere “*machine-written journalism*” (jornalismo escrito por máquina), ou ainda Carlson (2014) que descreve esta inovação como “*Robot Journalism*” (jornalismo robô) ou “*The Robot Reporter*” (repórter robô).

Em meio a tantas incertezas, é preciso reconhecer que, historicamente, as práticas jornalísticas sempre tiveram uma relação intrínseca e foram moldadas pelos avanços tecnológicos. A evolução dos meios de comunicação, comumente associada à ideia de redução das distâncias entre as pessoas, só se viabilizou com a invenção de aparatos técnicos como a prensa de Gutenberg, o telégrafo, o rádio, a televisão, a internet, entre tantos outros. Como defende Linden (2016), a automação de tarefas jornalísticas já acontece há pelo menos quatro décadas e esta recente inovação talvez seja “*apenas mais um passo na história do jornalismo-maquínico*”. Não seriam os *softwares* NLG mais uma entre tantas outras tecnologias presentes nas redações?

São perguntas como estas e a curiosidade em compreender o cenário no qual o Jornalismo Automatizado é implementado que nos motivaram a conduzir esta pesquisa. Assim, estabelecemos como **objetivo geral** traçar as redes sociotécnicas mobilizadas pelo Jornalismo Automatizado em um esforço de abrir as caixas-pretas e revelar os agenciamentos de atores humanos - jornalistas, programadores, empreendedores – e não-humanos – computadores, *softwares*, bancos de dados vinculados a esta inovação tecnológica.

Como **objetivos específicos**, esta pesquisa se propõe a:

- compreender as controvérsias e as incertezas desencadeadas pelo Jornalismo Automatizado em torno de uma “suposta” autonomia tecnológica de “robôs” redatores de notícias;
- apresentar experiências e veículos jornalísticos que adotaram *algoritmos* para a produção automatizada de notícias a partir da cartografia de três estudos de caso;
- evidenciar as negociações e tensões que emergem da relação entre jornalismo e tecnologia com o crescente processo de digitalização das práticas jornalísticas.

Como fundamentação teórica, nos inspiramos no STS - *Science and Technology Studies* em inglês – e na Teoria Ator-Rede em busca de um ponto de vista simétrico e não-antropocêntrico que nos auxilie a observar as redes sociotécnicas vinculadas a esta inovação, com um olhar atento para os agenciamentos e para o “programa de ação” que origina as notícias automatizadas. Tentamos desenvolver nossa argumentação baseada no discurso de “desconstrução” (AHVA e STEENSEN, 2017), com a intenção de não privilegiar nem uma perspectiva pessimista, nem determinista, e uma das influências teóricas foram duas pesquisas de STS sobre jornalismo digital conduzidas por Boczkowski (2004, 2010).

Nos aproximamos da discussão em torno da autonomia tecnológica (WINNER, 1977) e da mediação técnica (LATOURET, 1994b) em busca de referências que nos ajudassem a refletir sobre a crença de alguns pesquisadores que afirmam que as notícias automatizadas são produzidas sem a intervenção humana. Em um cenário repleto de incertezas, estes conceitos nos ajudam a compreender que esta é uma visão simplificada que não reconhece que, por trás destes textos, existe uma complexa rede sociotécnica de jornalistas, programadores, analistas de dados, empreendedores, pesquisadores, *softwares*, algoritmos e bancos de dados. Esta perspectiva nos auxilia ainda a refletir sobre os valores éticos e os contextos sociais e políticos nos quais esta tecnologia é desenvolvida e implementada.

Para observar os atores, a formação de redes e as associações vinculadas ao Jornalismo Automatizado, somos influenciados pelas lentes de observação sugeridas pela Cartografia de Controvérsias (VENTURINI, 2010, 2012) como um modo de olhar para esta inovação, o que fundamenta a nossa metodologia.

Assim, esta pesquisa busca desenvolver um olhar crítico e reflexivo que contextualize a produção automatizada de notícias em um cenário onde há um crescente uso de *softwares* e algoritmos que modificam as práticas profissionais e o consumo de informação. Para muitas pessoas, estes atores não-humanos continuam invisíveis, mas eles estão agindo cada vez mais e vieram para ficar. Sob o risco de encaixapretar as redes sociotécnicas, ignorá-los não é uma opção.

1.1 Sobre a estrutura desta dissertação

Esta dissertação está dividida em três partes. No próximo capítulo, apresentamos os fundamentos teóricos e metodológicos que nos guiaram neste processo, que se estabelecem principalmente no STS e na Teoria Ator-Rede, em busca de um olhar que se atente para as materialidades e para os contextos sociais e políticos em que o Jornalismo Automatizado é inserido no cotidiano das redações. Para sustentar a argumentação de que os processos algorítmicos que dão origem às notícias automatizadas são resultado de uma complexa rede sociotécnica que não elimina a dimensão humana e nem está limitada apenas à etapa de programação, dois conceitos teóricos foram fundamentais nesta trajetória: a autonomia tecnológica (WINNER, 1977) e a mediação técnica (LATOURETTE, 1994^b). Como perspectiva metodológica, esta pesquisa busca inspiração nas lentes de observação sugeridas pela Cartografia de Controvérsias (VENTURINI, 2010, 2012) para mapear pesquisas e estudos de caso sobre o Jornalismo Automatizado. Esta metodologia nos auxilia a identificar os atores humanos e não-humanos vinculados a esta inovação, a observar como eles se articulam em rede e a destacar alguns argumentos e declarações que emergem deste debate.

No capítulo 3, apresentamos um primeiro movimento cartográfico que descreve pesquisas e o contexto acadêmico onde se localiza o Jornalismo Automatizado, que argumentamos ser um campo de estudos em formação. Este capítulo começa com uma contextualização histórica que relata algumas experiências de produção automatizada de previsões do tempo, a partir da década de 1970, que foram a primeira aplicação desta tecnologia ao jornalismo. Explicamos também como funcionam os *softwares* NLG, e nos aproximamos do Jornalismo Computacional e de outras pesquisas que refletem sobre o uso dos algoritmos e do *big data* pela profissão. Por fim, descrevemos os resultados de algumas pesquisas acadêmicas específicas sobre a área que, desde 2012, se atentam a analisar como jornalistas e leitores reagem às notícias automatizadas.

No capítulo 4, apresentamos o segundo movimento cartográfico que se dedica a descrever três estudos de caso onde o Jornalismo Automatizado é implementado efetivamente nas redações, todos eles localizados nos Estados Unidos.

O primeiro estudo de caso apresenta a trajetória da *startup Narrative Science*, desde o surgimento do protótipo *StatsMonkey* em uma disciplina de pós-graduação na *Northwestern University* em 2009, que deu origem à empresa. A *Narrative Science* foi uma das grandes responsáveis pela divulgação desta tecnologia por conta de uma polêmica afirmação dada por um de seus fundadores, o professor Kristian Hammond, que declarou ao *The New York Times* em 2011 que “em cinco anos, um programa de computador vai ganhar o Pulitzer, e eu

serei condenado se não for a nossa tecnologia”⁴ (LOHR, 2011). Logo depois, a *startup* firmou parcerias com vários veículos jornalísticos, entre eles a *Forbes* e a *ProPublica*, mas a partir de 2016 eles decidiram mudar de estratégia mercadológica e abandonaram o jornalismo, focando na aplicação desta tecnologia para a redação de relatórios de *Business Intelligence*.

O segundo estudo de caso descreve as experiências em torno de outra *startup* denominada *Automated Insights*. A cartografia começa em 2010, quando a empresa ainda se chamava *StatSheet* e adotava esta tecnologia para alimentar uma rede de 345 sites que abrangia a publicação de 15 mil notícias por mês com a cobertura dos times de basquete da liga universitária dos Estados Unidos. A partir de 2014, a *Automated Insights* estabelece uma parceria com a agência de notícias *Associated Press* para a publicação de balancetes financeiros de empresas listadas na bolsa, que é sucedida pela criação do cargo de “editor de notícias automatizadas” ocupado pelo jornalista Justin Myers.

O terceiro estudo de caso relata dois projetos desenvolvidos pela equipe de jornalismo de dados do *Los Angeles Times*. Lançado em 2010, o primeiro projeto se chama *The Homicide Report* e adotou o Jornalismo Automatizado com o intuito de publicar uma nota para cada homicídio reportado pela Delegacia de Los Angeles, com informações básicas de apuração como nome da vítima, causa da morte, localização do incidente. No ar desde então, este projeto tem como premissa buscar uma forma igualitária de tratar os homicídios, dando também visibilidade para os casos mais frequentes envolvendo jovens negros assassinados nas regiões mais pobres da cidade. O segundo projeto se chama *Quakebot*, e atraiu a atenção da mídia em 2014 quando foi o primeiro a reportar um terremoto ocorrido em Los Angeles, apenas alguns minutos após o incidente. Nesta cartografia, o *Los Angeles Times* se destaca por ter desenvolvido dois projetos de Jornalismo Automatizado com sua equipe interna de jornalistas que também atuaram como programadores, e não a partir da parceria com uma *startup* como nos exemplos anteriores.

No final desta dissertação, refletimos sobre o conteúdo que emerge desta cartografia dialogando com a fundamentação teórica desta pesquisa, em uma análise que busca compreender a dimensão humana e a complexidade das redes sociotécnicas que atuam na produção automatizada de notícias. Argumentamos que existe um novo ecossistema em formação onde cada vez mais jornalistas trabalham em equipes multidisciplinares, ao lado de engenheiros, programadores, analistas de dados, envolvidos ainda com universidades, pesquisadores, *startups* e empreendedores.

⁴ Texto original: “In five years, a computer program will win a Pulitzer Prize — and I’ll be damned if it’s not our technology.” (LOHR, 2011)

2 Jornalismo pelo STS e pela TAR: interfaces e metodologias

Esta pesquisa se inspira nos trabalhos de STS e da Teoria Ator-Rede (TAR) como um modo de olhar para o Jornalismo Automatizado que esteja atento aos contextos sociais e políticos do desenvolvimento e implementação desta tecnologia, e que não limite nossa observação a uma análise purificada em torno das capacidades de produção de notícias por *softwares* NLG. Neste capítulo, apresentamos as referências teórico-metodológicas que nos guiaram nesta pesquisa, subdivididas em cinco subitens.

Na primeira parte, introduzimos as abordagens de STS e da TAR e destacamos alguns conceitos que foram essenciais para a trajetória desta pesquisa. No segundo subitem, descrevemos como o STS e a TAR dialogam com os estudos de jornalismo, onde apresentamos especificamente duas pesquisas sobre jornalismo digital realizadas por Boczkowski (2004, 2010) que nos ajudaram a compreender o contexto no qual o Jornalismo Automatizado é inserido nas redações jornalísticas. A terceira parte é dedicada à discussão de dois conceitos-chave que nos auxiliaram a desconstruir a imagem de que as notícias automatizadas são produzidas sem a intervenção humana: a autonomia tecnológica (WINNER, 1977) e a mediação técnica (LATOUR, 1994^b).

Com um enfoque mais metodológico, o quarto subitem discute a importância dos estudos de casos nas pesquisas em STS e apresenta a cartografia de controvérsias como um modo de olhar para o Jornalismo Automatizado. Em diálogo com Venturini (2010, 2012), conduzimos a pesquisa em busca dos atores, das literaturas, das declarações para compreender como são estabelecidas as associações e as redes sociotécnicas em torno desta inovação tecnológica. No último subitem, descrevemos como cartografamos o Jornalismo Automatizado nesta pesquisa, foco dos dois capítulos seguintes desta dissertação.

2.1 *Science and Technology Studies* (STS) e Teoria Ator-Rede (TAR)

Esta pesquisa busca analisar o Jornalismo Automatizado a partir de um enquadramento teórico interdisciplinar inspirado pelo STS⁵ - *Science and Technology Studies*

⁵ Existem dois significados para a sigla STS, que pode ser traduzida tanto como *Science and Technology Studies* quanto como *Science, Technology & Society*. Inicialmente, estes dois significados representavam uma divisão deste campo, sendo o primeiro STS um grupo filosófico mais radical preocupado em compreender a ciência e a tecnologia como atividades discursivas, sociais e materiais, e o segundo STS dedicado a refletir sobre as questões sociais que emergem deste desenvolvimento. Mas Sismondo (2010) defende que a expansão dos estudos de STS ao longo dos anos acabou por tornar as diferenças entre estes dois grupos cada vez menores. (SISMONDO, 2010, p. viii) Neste trabalho, optamos por utilizar a sigla em inglês, mas vale pontuar que, na América Latina, este campo de estudos é apropriado por pelo menos duas diferentes perspectivas, denominadas Ciência, Tecnologia e Sociedade (CTS) e Estudos Sociais de Ciência e Tecnologia (ESCT).

em inglês –, um programa que surgiu na década de 1970 quando um grupo de pesquisadores propôs uma nova forma de olhar para a ciência e para a tecnologia como disciplinas acopladas. Este campo de estudos questiona as abordagens clássicas da filosofia da ciência – como o positivismo lógico, o falsificacionismo e o realismo – que conferem à tecnologia um papel secundário e inferior ao conhecimento científico, e se sustentam em análises que posicionam a ciência como a “descoberta de verdades” e a tecnologia como a “aplicação destas verdades”. (SISMONDO, 2010)

Para os fundadores do STS, como Trevor Pinch, Wiebe Bijker, Thomas P. Hughes e Michel Callon, estas abordagens clássicas são fundamentadas em análises superficiais, simplificadas e lineares que desconsideram os contextos históricos e sociais nos quais as tecnologias são desenvolvidas. Uma das principais críticas desta perspectiva é acerca dos dualismos que fundamentam a mentalidade moderna, como ciência x tecnologia, natureza x sociedade, puro x aplicado. Ao criticarem estas abordagens que separam o conhecimento da prática, estes pesquisadores defendem um olhar social para o desenvolvimento tanto da ciência quanto da tecnologia, levando em conta os contextos e os indivíduos envolvidos nestes processos, que não podem ser analisados de forma separada e linear.

A ciência e a tecnologia se tornaram misturadas. A tecnologia moderna envolve cientistas que “fazem” tecnologia e tecnólogos que trabalham como cientistas (...) A antiga visão de que as ciências básicas geram todo o conhecimento que será aplicado pelos tecnólogos simplesmente não ajudará na compreensão da tecnologia contemporânea.⁶ (LAYTON, 1977 em PINCH e BIJKER, 1987, p. 20)

Para o STS, a relação entre ciência e tecnologia não pode ser reduzida a uma mera tradução do conhecimento científico que dá origem a objetos a partir de processos puramente mecânicos. (SISMONDO, 2010, p. 11) Ao desenvolver um ponto de vista que considera tanto a ciência quanto a tecnologia como culturas socialmente construídas, ao longo dos anos o STS desenvolveu várias abordagens que buscam analisar as mudanças tecnológicas a partir da observação de cientistas e engenheiros, e o modo como eles usam o mundo material em seus trabalhos para criarem redes e construir estruturas estáveis que mobilizam uma enorme quantidade de recursos ao longo do percurso. (Idem)

Uma das abordagens iniciais do STS recebeu o nome de *The Social Construction of Technology* (SCOT), um modelo multidirecional criado por Trevor Pinch e Wiebe Bijker (1987) que sustenta que as etapas “bem-sucedidas” do desenvolvimento de uma tecnologia não são os únicos caminhos possíveis para estudá-los, se opondo ao modelo estático e unidirecional

⁶ Texto original: “Science and technology have become intermixed. Modern technology involves scientists who ‘do’ technology and technologists who function as scientists (...) The old view that basic sciences generate all the knowledge which technologists then apply will simply not help in understanding contemporary technology.” (LAYTON, 1977 in PINCH and BIJKER, 1987, p. 20)

que considera a tecnologia como a aplicação do conhecimento científico. Esta metodologia propõe observar os processos que dão origem a uma nova tecnologia, identificando os relevantes 'grupos sociais' que se mobilizam e os diferentes caminhos possíveis que surgem ao longo deste desenvolvimento. Assim, ela demonstraria a existência de uma “flexibilidade interpretativa” nos processos que criam novos artefatos tecnológicos, com diferentes mecanismos de “enclausuramento” e “estabilização”.

Foi na mesma época em que surgiam as primeiras discussões em torno do STS que Michel Callon (1987) utilizou pela primeira vez a palavra ator rede – *actor network* em inglês e ainda sem o hífen – para se referir à ideia de “associações heterogêneas” com o intuito de distinguir esta terminologia dos “tradicionais atores da sociologia” que geralmente excluem os não-humanos de suas análises. “Um ator rede é simultaneamente um ator cuja atividade é interligar elementos heterogêneos e uma rede que é capaz de redefinir e transformar o que a constitui.”⁷ (CALLON, 1987, p. 93) Este conceito dá origem, posteriormente, à Teoria Ator-Rede (TAR), uma outra abordagem teórico-metodológica também vinculada ao STS desenvolvida a partir da década de 1980 pelos pesquisadores Michel Callon, Bruno Latour e John Law, entre outros. Também denominada de “sociologia das associações” (LATOUR, 1991) ou de “sociologia das translações” (LAW, 1992, p. 380), a Teoria Ator-Rede (TAR) propõe um olhar transversal e transdisciplinar para a ciência e para a tecnologia voltada para a formação de redes a partir das associações entre atores humanos e não-humanos. As principais referências destes pesquisadores são Michel Serres e o conceito de tradução, Gilles Deleuze e Felix Guattari e a noção de rizoma, Michel Foucault e a concepção de dispositivo, e David Bloor, que introduz o princípio da simetria. (FREIRE, 2006, p. 47)

Para a Teoria Ator-Rede (TAR), os mecanismos de “enclausuramento” e “estabilização” introduzidos pelo SCOT ganham o nome de caixa-preta, expressão “usada em cibernética sempre que uma máquina ou um conjunto de comandos se revela complexo demais” (LATOUR, 2000, p. 14). As caixas-pretas escondem em suas estruturas “certinhas” e “frias” várias incertezas e controvérsias que surgiram e obrigaram cientistas e engenheiros a tomarem decisões ao longo do desenvolvimento de uma tecnologia. O encaixapretamento “transforma um complexo objeto ou tecnologia, como uma porta, em uma simples ferramenta que pode ser usada na prática. E apesar de sua interna complexidade, ela pode ser tratada como uma unidade singular.”⁸ (BUEGER e STOCKBRUEGGER, 2015, p. 9)

⁷ Texto original: “An actor network is simultaneously an actor whose activity is networking heterogeneous elements and a network that is able to redefine and transform what it is made of.” (CALLON, 1987 p.93)

⁸ Texto original: “transforms a complex object or technology, like a door, into a simple tool that can be used in practice. And despite its internal complexity, it can be treated as a single unit.” (BUEGER e STOCKBRUEGGER, 2015, p. 9)

A TAR se fundamenta em dois conceitos-chave: o ator e a rede, sendo que “as duas partes são essenciais, daí o hífen” (LATOUR, 2012, p. 260). A rede é o espaço-tempo da ação, o movimento associativo entre elementos onde surge o social (CALLON, 1987). Não é uma entidade, mas um conceito. A rede é o que se forma a partir das conexões que um ator realiza com outros; é uma série de associações que permitem compreender a continuidade da ação. (LATOUR, 2013, p. 33) Atores – também denominados de actantes ou mediadores – são agentes que “transformam, traduzem, distorcem e modificam o significado ou os elementos que supostamente veiculam” (LATOUR, 2012, p. 65).

Agenciar é fazer algo ou alguém agir, é colocar em contato dois elementos para uma possível relação. A ação surge como resultado de processos de tradução/translação, pequenas transformações que se tornam possíveis a partir da conexão provisória de pessoas e objetos que são reconhecidos não por suas qualidades individuais, mas por fazerem parte de uma rede, de um coletivo que dá origem aos híbridos. Esta perspectiva teórica defende que a sociedade é composta de uma rede heterogênea não só de pessoas, mas também de máquinas, animais, objetos, textos, dinheiro, arquiteturas, e uma infinidade de outros não-humanos que se destacam no momento de descrição de uma rede. Um ator nunca age sozinho nem é possível afirmar a quem pertence a ação, pois a “ação não é propriedade somente dos humanos, mas uma associação de actantes” (LATOUR, 1994^b, p. 35)

No livro “Jamais Fomos Modernos”, Latour (1994^a) critica a mentalidade moderna fundamentada em uma lógica linear e na separação entre natureza e sociedade, que continua condicionando o pensamento pós-modernista e não leva em conta o papel dos não-humanos e dos híbridos na sociedade. Ao afirmar que nunca fomos modernos e nem somos pós-modernos, Latour (1994^a) defende uma perspectiva que concebe a sociedade como uma entidade plural, coletiva, onde é necessário assumir um ponto de vista simétrico e não-antrópocêntrico que considere os agenciamentos entre atores humanos e não-humanos em suas relações de mediação e tradução/translação.

Ao questionar os dualismos sobre os quais se fundamenta a ideia de Modernidade e propor uma perspectiva que confere responsabilidade tanto para os homens quanto para os objetos, a TAR se opõe tanto ao materialismo que defende o determinismo tecnológico, quanto ao moralismo e sua visão antropocêntrica do determinismo humano sobre a técnica. (LATOUR, 1994^b)

Neste contexto, a TAR assume como um de seus pilares o princípio da simetria que defende um olhar antropológico que não conceda privilégios a nenhum dos lados, nem à natureza, nem à sociedade. Nas palavras de Latour (1994^a, p. 95), “é preciso que a antropologia absorva aquilo que Michel Callon chama de princípio da simetria generalizada: o antropólogo deve estar situado no ponto médio, de onde pode acompanhar, ao mesmo tempo,

a atribuição de propriedades não-humanas e de propriedades humanas”, o que evidencia a perspectiva não-antropocêntrica assumida pela teoria.

No meio, onde supostamente nada acontece, quase tudo está presente. E nas extremidades, onde reside, segundo os modernos, a origem de todas as forças, a natureza e a sociedade, a universalidade e a localidade, não há nada além de instâncias purificadas que servem de garantias constitucionais para o conjunto. (LATOUR, 1994^a, p. 121)

Esta busca por uma abordagem simétrica imprime na TAR um desejo de “representar atores humanos e não-humanos, e tratá-los nos mesmos termos relacionais, em uma forma de criar análises completas que não descriminem nenhuma parte das ecologias de fatos científicos e objetos tecnológicos”⁹ (SISMONDO, 2010, p. 87). Este princípio exige que o pesquisador abandone algumas crenças e incorpore um olhar que não crie delimitações predeterminadas sobre ciência e tecnologia.

Para Latour, ao seguir os passos dos cientistas, o pesquisador não precisa se preocupar em classificar se uma associação é “social” ou “científica”, por exemplo, e sua atenção deve estar focada no movimento dos atores que ele está seguindo, cuidadoso em observar, ao longo do processo, quais associações são fortes e estáveis e quais são fracas e correm o risco de se soltar. Ao longo das observações, o pesquisador deve permanecer tão indeciso quanto os vários atores que segue e, sempre que for definida uma divisão entre interior e exterior, ele deve acompanhar simultaneamente os dois lados e criar uma lista – muitas vezes longa e heterogênea – daqueles que realmente trabalham.

Law (1992) ainda destaca que uma das preocupações da TAR é compreender as mecânicas do poder e das organizações. Para isso, um dos preceitos básicos para o pesquisador é não assumir como ponto de partida em sua análise que, de um lado, existe um sistema macrossocial, e de outro várias peças soltas derivadas de um microssocial, pois assim ele se fecha logo no início para as questões mais interessantes e que dão origem às relações de poder. (Idem, p. 380)

Para observar estas redes dinâmicas e heterogêneas compostas por atores humanos e não-humanos, a TAR defende que o pesquisador observe como os fatos são construídos e esteja atento a identificar os “dispositivos de inscrição” (LATOUR, 2000, p. 112), definidos como qualquer elemento que o permita acompanhar o movimento dos atores, tais como instrumentos, gráficos, tabelas, mapas. Uma inscrição “descreve uma relação estável entre dois atores heterogêneos” e pode transformar uma complexa tecnologia em uma caixa-preta

⁹ Texto original: “Representing both human and non-human actors, and treating them in the same relational terms, is one way of prompting full analyses, analyses that do not discriminate against any part of the ecologies of scientific facts and technological objects.” (SISMONDO, 2010, p. 87)

que age como se fosse um objeto único e não uma rede. (BUEGER e STOCKBRUEGGER, 2015, p.10).

Estas redes compostas por humanos e não-humanos são descritas por Latour (1994^a) como uma “cadeia de mediações”, onde não é possível definir de forma precisa o que torna algo humano ou técnico. Estamos envolvidos em alianças em constantes modificações, negociações e rearranjos, representações e delegações. Deste modo, o autor argumenta que o problema reside no fato de, seguindo a mentalidade moderna, os sociólogos terem se tornado inimigos dos tecnólogos, quando na verdade eles deveriam ser irmãos. Para Latour, a separação entre o social e os objetos nos impede de compreender o mundo onde vivemos.

2.2 O jornalismo a partir da perspectiva do STS e da Teoria Ator-Rede

Desde o início dos anos 2000, tanto o STS quanto a Teoria Ator-Rede têm influenciado vários estudos sobre jornalismo, em um movimento que é denominado por Benson (2017) como uma “nova era descritivista”. Nesta pesquisa, este campo teórico-metodológico serve de inspiração para direcionar o olhar para os contextos sociais e políticos nos quais são aplicados os *softwares* NLG no jornalismo, evidenciando os atores e as redes sociotécnicas mobilizadas ao longo deste processo.

Esta escolha é ainda motivada pela vontade de buscar um ponto de vista que não analise as notícias automatizadas a partir de uma concepção moderna e linear do jornalismo que, como apontado por Turner (2005), constantemente separa o humano do mecânico, “as notícias dos jornalistas, os repórteres da audiência, a imprensa da política”. Ao analisar as práticas jornalísticas e as organizações sociais vinculadas a ela a partir da perspectiva da TAR, Turner (2005) reconhece a emergência dos híbridos, uma característica cada vez mais frequente no jornalismo digital onde não é mais possível analisar, sob um ponto de vista antropocêntrico e modernista, o jornalismo a partir da ação somente dos três principais atores humanos: fontes, jornalistas e audiência. (TURNER, 2005, p. 322) Há mais de uma década, portanto, o autor já chamava a atenção para o fato das práticas jornalísticas – e da comunicação como um todo - sempre terem sido moldadas pelas tecnologias, o que corrobora a importância de se refletir sobre a profissão a partir da perspectiva do STS e da TAR, levando em consideração a ação dos não-humanos.

Boczkowski e Lievrouw (2008) criticam a forma como duas das principais tradições de pesquisas em *Media Studies* lidam com a tecnologia ou como extensões do homem, destacando suas capacidades técnicas, ou orientada por uma abordagem behaviorista que realça a complexidade das tecnologias contemporâneas de mídia e sua dependência com a computação. (BOCZKOWSKI, LIEVROUW, 2008, p. 952) Os autores argumentam que ambas

perspectivas se assemelham pelo foco nas capacidades técnicas da tecnologia, e não se atêm ao fato de que as materialidades devem ser situadas em um contexto social, cultural, temporal, político, econômico. Lievrouw (2014) defende que, para criar estudos que estejam na interseção entre o STS e a comunicação, os pesquisadores devem adotar metodologias que explicitem a conexão entre ferramentas tecnológicas, ações humanas e formações socioculturais. (LIEVROUW, 2014 p. 44)

Ahva e Steensen (2017) pontuam ainda que as primeiras pesquisas que refletiram sobre os impactos da tecnologia digital sobre o jornalismo enfatizavam uma perspectiva determinista e utópica, associando estes novos artefatos técnicos a um discurso revolucionário. Em contraposição a esta abordagem, eles citam o surgimento de um contra discurso estabelecido na ideia de evolução, enfatizando por exemplo a resiliência das práticas jornalísticas que não são drasticamente modificadas com a digitalização, e sim moldadas a partir das práticas anteriormente estabelecidas. Ahva e Steensen (2017) sugerem a construção de uma perspectiva interdisciplinar para as pesquisas sobre jornalismo digital fundamentada em quatro campos do conhecimento – sociologia, ciência política, estudos culturais e perspectivas sócio-materiais como o STS – e estabelecidos na ideia de um discurso de “desconstrução”.

Ahva e Steensen (2017) ressaltam ainda as contribuições que as perspectivas sócio-materiais ligadas à tecnologia – como a teoria da inovação, o SCOT, a antropologia da tecnologia e o STS – trouxeram para as pesquisas em jornalismo *online*, destacando o papel da Teoria Ator-Rede adotada por alguns estudos para refletir sobre o papel desempenhado pela materialidade nos processos jornalísticos e na formação de redes em torno de uma notícia.

Com uma crescente ênfase na materialidade e na tecnologia, o conceito de rede se popularizou, com foco nas possibilidades proporcionadas pelo ambiente digital de rastrear as associações entre vários atores (...) Além disso, a promessa do *big data* para as pesquisas e práticas jornalísticas provocaram discussões entre os estudiosos, e alguns passos foram tomados de uma mera investigação empírica e tecnológica para uma compreensão mais holística – abrangendo aspectos de epistemologia, expertise, ética e economia – sobre o *big data* no contexto do jornalismo digital.¹⁰ (AHVA, STEENSEN, 2017, p. 31)

Boczkowski (2004, 2010) é um dos primeiros pesquisadores a adotar o STS para conduzir duas pesquisas sobre jornalismo digital e, ainda que não tratem diretamente da

¹⁰ Texto original: “In the increased emphasis in materiality and technology, the concept of the network has become popular, focusing on the ways the digital environment has provided possibilities for tracing the associations between various actors (...) Furthermore, the promises of ‘big data’ for journalism research and practice have evoked discussions among scholars, and steps have been taken from merely technological or empirical research toward a more holistic understanding – embracing aspects of epistemology, expertise, ethics, and economics – of big data in the context of digital journalism.” (AHVA, STEENSEN, 2017, p. 31)

produção automatizada de notícias, suas pesquisas trazem alguns *insights* interessantes sobre as dinâmicas sócio-materiais dos meios de comunicação e servem de inspiração para esta pesquisa. Se opondo ao discurso que comumente associa as novas mídias aos efeitos revolucionários relacionados à tecnologia e à internet, Boczkowski (2004) sugere o conceito mídias emergentes – *emerging media* em inglês – para defender a ideia de que as inovações digitais acontecem de maneira gradual e moldadas pelas infraestruturas já existentes das mídias impressas e eletrônicas, enfatizando questões históricas, locais e processuais associadas às novas tecnologias. (Idem, p.12) Inspirado em Latour (1987), Boczkowski (2004) destaca a importância de “seguir os atores” para refletir sobre a “reconstrução das notícias em ambiente online”¹¹ (Idem, p. 183), destacando que uma das conclusões de sua pesquisa é a de que as notícias produzidas para as mídias digitais mobilizam um grupo maior e mais heterogêneo de atores, se comparado com as equipes envolvidas na produção de notícias para as mídias impressa e eletrônica.

É importante pontuar que as notícias automatizadas são um fenômeno situado em um contexto de digitalização das práticas jornalísticas e, semelhante às observações de Boczkowski (2004), elas se tornam possíveis a partir da ação de um grupo heterogêneo de atores onde jornalistas, cada vez mais, trabalham em equipes multidisciplinares incluindo profissionais de tecnologia, engenheiros, analistas de dados, empreendedores. Ao considerar o uso de *softwares* NLG por veículos jornalísticos como uma mídia emergente, buscaremos desenvolver um ponto de vista mais amplo que não olha esta inovação de forma isolada e purificada, e sim tenta analisá-la sob um espectro mais amplo que se preocupa com o contexto no qual esta tecnologia se insere e é moldada por questões sociais, históricas, culturais, locais e de infraestrutura.

Ao analisar os padrões de inovação do jornalismo em três jornais diários norte-americanos, Boczkowski (2004) observa que o desenvolvimento de tecnologias se deu a partir de uma mentalidade que cultuava a impressão e se apropriava da ideia de que “o futuro seria uma versão melhorada, mas não radicalmente diferente, do presente”¹² (BOCZKOWSKI, 2004 p. 12). Este conservadorismo estabelece como um dos focos da digitalização a proteção dos privilégios dos jornais impressos, o que consequentemente reflete na criação de sites pouco inovadores. Adicionalmente, com os lucros advindos das assinaturas e da publicidade vinculadas à versão impressa, estas empresas também não se preocuparam no início dos anos 2000 em desenvolver um modelo de negócios sustentável para as suas versões *online*.

¹¹ Texto original: “reconstructing of news in the online environment” (BOCZKOWSKI, p. 183)

¹² Texto original: the future would be an improved, but not radically different, version of the present.” (BOCZKOWSKI, 2004 p. 12)

Este conservadorismo característico da profissão é um dos fatores que colabora para a disseminação de uma visão pessimista frente aos avanços tecnológicos, como acontece com o Jornalismo Automatizado. Os jornalistas são historicamente condicionados a reagirem negativamente à introdução de novos artefatos técnicos, e tendem a agir motivados pela manutenção de processos tradicionais e já cristalizados, demonstrando dificuldades para se adaptarem às inovações.

Boczkowski (2004) chama a atenção para a materialidade como um dos elementos fundamentais que molda os processos adotados por uma redação *online*, onde os artefatos técnicos influenciam as práticas adotadas pelos jornalistas, os assuntos que serão abordados, a forma como serão narrados e para qual público serão direcionados. O autor chama a atenção para o fato de que, em muitas redações *online*, a produção de conteúdos se limitaria à reprodução de notícias e reportagens publicadas pela versão impressa, e as equipes teriam pouca autonomia para produzirem conteúdos próprios explorando as características de uma publicação digital. (Idem, p. 171-172) Boczkowski (2004) discorda da separação entre *online* e *offline* da maioria das pesquisas que discutem a mediação computacional na comunicação, sugerindo que as características *offline* do conteúdo e dos artefatos influenciam a experiência online dos usuários.

Em um segundo estudo, Boczkowski (2010) parte da premissa de que o jornalismo *online* introduz um novo padrão temporal de consumo de notícias, fenômeno denominado pelo autor de “notícias no trabalho” – *news at work* em inglês. Se o jornal impresso é produzido preferencialmente para ser consumido durante o café da manhã, antes do trabalho, e o jornal televisivo é preferencialmente consumido à noite, no “horário nobre”, o mesmo não acontece com as notícias online, onde sistemas de monitoramento de audiência como o *Google Analytics* apontam para um consumo maior de notícias de segunda a sexta entre 9h e 18h, horário em que, normalmente, as pessoas estão trabalhando. Instigado por este fenômeno, Boczkowski (2010) desenvolve uma pesquisa etnográfica e interdisciplinar observando as redações online de dois veículos argentinos – o *Clarín* e o *La Nación* – e observa como este padrão de audiência nos sites moldou transformações editoriais, de estrutura organizacional e de design gráfico, como por exemplo o aumento na frequência de atualizações e no volume de notícias publicadas ao longo do dia.

Boczkowski (2010) argumenta que a mudança no padrão de consumo de notícias ao longo do dia modifica a temporalidade com a qual os sites jornalísticos são atualizados, intensificando o monitoramento de sites concorrentes que geram, consequentemente, um aumento no grau de imitação e uma queda na diversidade do conteúdo publicado. Esta crescente imitação e homogeneização das notícias é denominada por ele de “espiral da mesmice” – *spiral of sameness* em inglês – em uma paráfrase à famosa teoria da “espiral do

silêncio” de Noelle-Neumann (1993). “Na era da abundância de informação, o que a maioria dos leitores obtém é mais do mesmo.”¹³ (BOCZKOWSKI, 2010, p. 6) Boczkowski (2010) cita ainda o famoso texto de Walter Benjamin (1935) intitulado “A obra de arte na era de sua reprodutibilidade técnica” para evocar a preocupação do sociólogo quanto à perda da aura nas obras de arte, que poderia neste caso ser aplicada às notícias. (Idem, p. 172)

A dinâmica de imitação nos veículos jornalísticos começa no processo de produção das notícias, se materializa nos textos publicados, modifica os hábitos de consumo dos leitores, e gera ainda implicações políticas e culturais. Boczkowski (2010) pontua que esta crescente homogeneização não agrada nem os jornalistas nem os leitores, mas ambos não possuem autonomia e autoridade suficiente para alterarem esta lógica de produção, que implica ainda em um enfraquecimento no papel de vigilância – *watchdog* em inglês - tradicionalmente exercido pelas organizações jornalísticas, que conseqüentemente pode afetar o equilíbrio do poder na sociedade (Idem, p. 6). Entre algumas conseqüências geradas por este fenômeno, o autor prevê o aumento da publicação de notícias genéricas pela mídia tradicional, em parte fornecidas por agências e por terceiros, e a redução de iniciativas independentes que desempenhem a função de vigilância do jornalismo.

Inspirada por este estudo de Boczkowski (2010), é possível associar este fenômeno denominado pelo autor de “espiral da mesmice” ao contexto da produção automatizada de notícias. Se o crescente grau de imitação e homogeneização das notícias é identificado anteriormente no jornalismo digital, ele adquire uma proporção ainda mais ampla quando *softwares* NLG são inseridos nas redações jornalísticas. Estes sistemas são configurados para produzirem e publicarem um grande volume de narrativas a partir de templates formulaicos ligados a bancos de dados estatísticos estruturados. Na prática, o que ocorre é a publicação de milhares de notícias com estruturas simples e repetitivas, semelhantes entre si e diferenciadas no sujeito em destaque e no processamento das variáveis numéricas. Por exemplo, enquanto no caso de jogos esportivos modificam-se os resultados e os times em evidência, no caso de balancetes financeiros variam-se o nome das empresas e os lucros e dividendos alcançados no trimestre.

O Jornalismo Automatizado eleva exponencialmente o grau de imitação e homogeneização dos conteúdos publicados, até o momento, exclusivamente em plataformas digitais, já que ainda não identificamos nenhum estudo de caso onde estas narrativas foram publicadas em veículos impressos ou eletrônicos. É importante pontuar ainda que tanto empresários de *startups* que desenvolvem *softwares* NLG quanto executivos de grandes

¹³ Texto original: In an age of information plenty, what most consumers get is more of the same.” (BOCZKOWSKI, 2010, p. 6)

empresas jornalísticas gostam de associar o uso de sistemas como este à “cauda longa” (ANDERSON, 2006). A produção e publicação destes conteúdos seria dirigida a um público restrito e específico seguindo uma lógica de nicho, e não de comunicação de massa, em um contexto onde novas tecnologias são inseridas, sustentadas pela promessa futura de uma crescente personalização tanto nas formas de consumo quanto no formato das narrativas.

2.3 Duas concepções de autonomia: Winner e Latour

Em busca de uma perspectiva que nos auxilie a abrir as caixas-pretas e olhar de forma mais específica para as redes sociotécnicas mobilizadas pelo Jornalismo Automatizado, em uma análise que não invisibilize a dimensão humana por trás desta inovação, dois conceitos nos parecem fundamentais neste processo: a autonomia¹⁴ tecnológica (WINNER, 1977) e a mediação técnica (LATOUR, 1994^b), como discutiremos a seguir.

Winner (1977) define autonomia tecnológica como “a crença de que de alguma forma a tecnologia está fora do controle e segue seu próprio curso, independente da direção humana”¹⁵ (WINNER, 1977, p. 13). Trata-se de uma ideia que se tornou uma obsessão na literatura dos séculos XIX e XX e que alimenta o medo da tecnologia não seguir mais aos propósitos humanos, tendo autonomia para agir por si mesma e escapar da dominação de seu criador. O autor pontua que esta mentalidade que acredita que a tecnologia está “fora do controle” tem natureza política e está enraizada em concepções antropocêntricas de poder e dominação da sociedade. De forma mais ampla, essa perspectiva estaria diretamente associada ao pensamento despótico de que o mestre tem controle sobre seus escravos, ou à expectativa que ecoa o pensamento de Aristóteles de que algo precisa ser escravizado para que o outro conquiste a sua emancipação. “O conceito de maestria e a metáfora mestre-escravo são as formas dominantes de se descrever a relação do homem com a natureza, bem como com as ferramentas tecnológicas.”¹⁶ (WINNER, 1977, p. 20)

¹⁴ É interessante observar que as palavras autonomia e automação/automatizado, recorrentemente citadas nesta pesquisa, compartilham do mesmo prefixo grego *auto*. De acordo com Partridge (1958), este prefixo é derivado da palavra “autêntico” e se refere à noção de “alguém que age por sua própria autoridade”. Seu significado está diretamente relacionado a termos como “eu”, “vida” e “da alma”, e seu uso pressupõe uma ideia de auto existência. Portanto, termos como automatizado, autônomo e autoridade, comumente associados às tecnologias, incorporam a ideia de ter uma vida própria atrelada ao prefixo *auto*. A diferença entre esses termos está condicionada a seus sufixos: o -matos de automatização vem da palavra “mente” e possui o significado de pensar ou “agir por si só”; e o -nomia de autonomia é derivado da palavra “ágil” e transmite a conotação de “autogoverno” e governança.

¹⁵ Texto original: “the belief that somehow technology has gotten out of control and follows its own course, independent of human direction.” (WINNER, 1977, p. 13)

¹⁶ Texto original: “the concept of mastery and the master-slave metaphor are the dominant ways of describing man’s relationship to nature, as well as to the implements of technology.” (WINNER, 1977, p. 20)

Assim, a tendência de associar o Jornalismo Automatizado à imagem de robôs ou conceituá-lo como processos algorítmicos que produzem notícias sem nenhuma intervenção humana pode ser analisada como uma concepção fundamentada na ideia de dominação e na existência de uma relação mestre-escravo entre o homem e a tecnologia. O jornalista, neste caso, desempenharia o papel de mestre que até então dominava os artefatos técnicos presentes no dia-a-dia de sua atuação profissional, mas ao se deparar com softwares que automatizam a redação de notícias, ele reage com apreensão e medo, temendo ter chegado o dia em que é decretada a alforria de seus escravos tecnológicos, que passariam então a atuar de forma autônoma e independente de qualquer ação humana.

Ao discutir as várias noções atreladas à autonomia tecnológica, Winner (1977) busca um ponto de vista médio que não pressuponha, seguindo a perspectiva dos “críticos radicais”, que o avanço tecnológico seja uma “força opressora que representa uma ameaça direta à liberdade humana”¹⁷ (Idem, p. 3), e nem que, seguindo a visão dos entusiastas, a tecnologia transformará de forma revolucionária o mundo. Ele defende a ideia de que o desenvolvimento tecnológico gera um crescente aumento na complexidade, o que exige das pessoas e dos sistemas sociais uma adaptação contínua. O autor argumenta que não há nada de novo nas mudanças acarretadas pelo avanço atual das técnicas, chamando a atenção para o fato de que, desde a Idade Média, a sociedade passa por um processo contínuo de mudanças sociotécnicas e é remodelada politicamente em resposta às inovações tecnológicas. “A tecnologia em suas várias manifestações é uma parte significativa do mundo humano. Suas estruturas, processos e alterações entram e se tornam parte das estruturas, processos e alterações da consciência humana, da sociedade e da política.”¹⁸ (Idem, p. 6)

Para Winner (1977), é preciso desmistificar algumas crenças do pensamento moderno ocidental que defendem, em tom de superioridade, que os homens sabem mais do que os objetos que eles mesmo desenvolveram, que as coisas estão sob o controle dos homens, que a tecnologia é essencialmente neutra e, portanto, os danos e benefícios por ela causados dependem da forma como os homens a utilizam. (Idem, p. 25) Convencionalmente, a sociedade foi condicionada a acreditar que a tecnologia é essencialmente neutra, nada mais do que apenas uma mera ferramenta a serviço do homem, uma percepção que desconsidera a complexidade das relações estabelecidas há séculos na “sociedade tecnológica”. Para Winner (1977), acreditar na neutralidade tecnológica é um mito, uma perspectiva que não consegue visualizar as relações de poder estabelecidas entre o homem e os artefatos técnicos

¹⁷ Texto original: “oppressive force that poses a direct threat to human freedom.” (Idem, p. 3)

¹⁸ Texto original: “Technology in its various manifestations is a significant part of the human world. Its structures, processes and alterations enter into and become part of the structures, processes, and alterations of human consciousness, society and politics.” (WINNER, 1977, p. 6)

que o rodeiam, que não se atenta para a resistência que muitos desses objetos manifestam orientados por objetivos ou padrões preconcebidos. A tecnologia não é capaz de agir sem a presença do homem, mas este não ocupa a posição de mestre como tradicionalmente muitos acreditam. Nesta relação, os homens tendem a obedecer às normas e aos requisitos, cada vez mais complexos, dos sistemas que alegadamente governam. (Idem, p. 29)

Winner (1977) sugere inclusive que a tecnologia é possivelmente mais autoritária até mesmo que o empregador, ou o patrão, exigindo que regras e procedimentos sejam cumpridos em situações onde os homens têm pouca ou nenhuma escolha a não ser a obediência estrita à técnica, em uma governança que flui dos artefatos técnicos em direção às pessoas e aos arranjos sociais, e não no sentido oposto. Assim, a tecnologia não é neutra e nem uma mera ferramenta que espera passivamente ser utilizada, mas um conjunto tecnopolítico que exige dos homens um comportamento padronizado e rotineiro.

Neste contexto, não podemos ser ingênuos em considerar que os *softwares* NLG são meras ferramentas neutras a serviço dos jornalistas. Estes sistemas estão inseridos em contextos sociais que modelam a sua capacidade de atuação, e fazem parte de uma extensa rede constituída por diversos profissionais e várias outras tecnologias que determinam o seu alcance e o limite de suas ações.

Semelhante à visão de rede da TAR, Winner (1977) argumenta que os “artefatos técnicos do século XX” são caracterizados pelo “enorme tamanho, complexas interconexões e uma interdependência sistêmica”¹⁹ (Idem, p. 200), onde uma parte é essencial para o trabalho das outras partes. “Componentes que foram desenvolvidos e operavam separadamente estão agora ligados em uma estrutura organizada. As redes resultantes representam um salto quântico sobre o poder e as capacidades de desempenho de sistemas menores e mais segmentados.”²⁰ (WINNER, 1977, p. 238)

Um dos objetivos de Winner (1977) é desmistificar a imagem negativa associada ao desenvolvimento tecnológico, em busca de um ponto de vista que aproxime a autonomia tecnológica da autonomia humana. Esta perspectiva propõe conceituá-la como “uma parte do nosso ser que foi transferida, transformada e separada das nossas necessidades vitais e inteligência criativa. Qualquer esforço para recuperar esta parte da vida humana deve ser vista, primeiramente, como impraticável e até absurda.”²¹ (WINNER, 1977, p. 333) Como no

¹⁹ Texto original: “enormous size, complex interconnection, and systemic interdependence.” (Ibidem, p. 200)

²⁰ Texto original: “Components that were developed and operated separately are now linked together to form organized wholes. The resulting networks represent a quantum jump over the power and performance capabilities of smaller, more segmental systems.” (WINNER, 1977, p. 238)

²¹ Texto original: “Autonomous technology is part of our being that has been transferred, transformed, and separated from living needs and creative intelligence. Any effort to reclaim this part of human life must at first seem impractical and even absurd.” (WINNER, 1977, p. 333)

mito de Frankenstein, o autor procura demonstrar que os artefatos técnicos reproduzem de forma grotesca uma força autônoma, em um reflexo da natureza humana, e ao mesmo tempo são incompletos, têm sua existência atrelada e dependente dos humanos, mas isso não significa que eles estão totalmente sob o nosso controle.

Neste livro, Winner (1977) nos suscita a estabelecer novos conceitos com relação ao papel da tecnologia na sociedade, que não podem ser reduzidos a uma mera relação de dominação entre mestre-escravo e nem a um olhar ingênuo que acredita na neutralidade destes objetos, e sim na busca de uma compreensão política que tenta observar quais forças de disputa e resistência são estabelecidas nas relações entre homens e objetos em uma “sociedade tecnológica”. Seguindo esta perspectiva, não há nenhum sentido imaginar uma situação onde os *softwares* NLG atuam de maneira isolada, autônoma e independente da ação humana nas redações jornalísticas. O uso de sistemas de produção automatizada de notícias só se viabiliza se inserido em um contexto social mais amplo onde se estabelecem relações de interdependência entre jornalistas e profissionais de outras áreas afins com as tecnologias que os cercam.

Esta reflexão em torno da autonomia tecnológica (WINNER, 1977) reverbera em um outro debate não menos importante para a busca, nesta pesquisa, por um embasamento teórico que nos auxilie a desconstruir a ideia que as notícias automatizadas são produzidas sem a intervenção humana, ou que limite a dimensão humana à etapa de programação de algoritmos. Trata-se da discussão em torno da mediação técnica (LATOURE, 1994^b), conceito vinculado à Teoria Ator-Rede e que possui seu significado atrelado à ideia de delegação.

Semelhante à argumentação de Winner (1977), Latour (1994^b) critica tanto a perspectiva moralista que associa a tecnologia à ideia de neutralidade dos objetos onde o humano se sobrepõe à técnica, quanto a perspectiva materialista que, de forma pessimista como defendido por Heidegger, acredita que a tecnologia determina a ação humana. Tomando como exemplo um homem e uma arma, para os moralistas a arma é um objeto neutro que só age na mão de um assassino, e para os materialistas a arma é o objeto que torna um homem assassino, mas para Latour (1994^b) um tiro só se concretiza a partir da associação destes dois agentes em um “programa de ação”, o que o leva a defender que “nenhuma ação sem mediação é possível”²² (Idem, p. 29) e, assim, “a responsabilidade pela ação deve ser compartilhada entre os vários actantes”²³ (Idem, p. 34). Em um agenciamento homem-objeto, ambos se modificam e influenciam o curso da ação, e destas ligações surgem os híbridos. Para Latour (1994^b),

²² Texto original: “No unmediated action is possible.” (LATOURE, 1994^b, p. 29)

²³ Texto original: “Responsibility for action must be shared among the various actants.” (Idem, p. 34)

objetos que existem apenas como objetos, acabados, que não fazem parte de uma vida coletiva, não são conhecidos, estão enterrados abaixo do solo. Objetos reais fazem sempre parte de instituições, vibrando em seu status misto de mediadores, mobilizando terras distantes e pessoas, prontas para se tornarem pessoas ou coisas, sem saber se são compostas de uma ou de muitas, de uma caixa-preta que esconde um labirinto de multidões.²⁴ (LATOUR, 1994^b, p. 46)

Latour (1994^b) propõe quatro significados para o que ele chama de mediação técnica, sendo o primeiro deles o “programa de ação”, a série de objetivos, passos e intenções que podem ser descritas em uma situação, também denominado de “translação” - inspirado em Michel Serres - para descrever a ligação que surge e que modifica dois actantes no curso de uma ação. (Idem, p. 32)

Como segundo significado para mediação, a ação é resultado de uma composição de forças e não é propriedade somente dos humanos, e sim de uma associação de actantes, em que deve prevalecer o princípio da simetria em substituição da dicotomia entre sujeito-objeto, uma classificação que dificulta a compreensão tanto da técnica quanto da sociedade.

O papel da mediação técnica em geral é difícil de ser medido com precisão porque a ação dos objetos passou por um processo de encaixapretamento que tornou opaco os agenciamentos entre sujeitos-objetos. Neste contexto, surge o terceiro significado de mediação, definido como “caixa-preta reversível”, onde ao examinar os actantes e as associações presentes em uma caixa-preta, nos deparamos com várias outras caixas-pretas repletas de novos atores, novas associações, uma situação paradoxal que nos leva a questionar qual o limite de reconstituição das redes sociotécnicas. São inúmeras entidades silenciosas, que agem pacificamente ao nosso redor, e não somos nem mesmo capazes de contá-las ou de definir se são objetos, um conjunto de associações, ou uma sequência de ações especializadas. (Idem, p. 37-38)

O quarto significado de mediação técnica, para Latour (1994^b) o mais importante deles, é a delegação e, para explicá-lo, o autor exemplifica com a descrição de um quebra-molas, um objeto moldado a partir da ação de operários, engenheiros, concreto, asfalto, tinta que posteriormente exerce seu “programa de ação”, obrigando os motoristas a reduzirem a velocidade de seu veículo em algum determinado ponto da via. Como o quebra-molas, vários outros objetos ocultam em sua rigidez material os vários humanos e não-humanos que, em um momento anterior ao fechamento das caixas-pretas, estiveram presentes e se associaram para tornar possível o “programa de ação”. Vivemos, assim, no meio de “delegações técnicas”

²⁴ Texto original: “Objects that exist simply as objects, finished, not part of a collective life, are unknown, buried under the soil. Real objects are always parts of institutions, trembling in their mixed status as mediators, mobilizing faraway lands and people, ready to become people or things not knowing If they are composed of one or of many, of a black box counting for one or of a labyrinth concealing multitudes.” (LATOUR, 1994^b, p. 46)

(Idem, p. 40), onde objetos nos obrigam a agir delegados por outros actantes que não vemos e nem conseguimos rastrear. “A ação técnica é uma forma de delegação que nos permite mobilizar, durante as interações, movimentos feitos em outros lugares, mais cedo, por outros actantes.”²⁵ (Ibidem, p. 52)

É importante pontuar que o ponto de vista de Latour (1994^b) sobre a mediação técnica, principalmente neste quarto significado que gira em torno da delegação, se difere da visão de Winner (1977) sobre a autonomia tecnológica. Enquanto o primeiro defende que uma ação técnica é resultado da mobilização de vários outros actantes que no passado foi encaixapretada em um objeto, por exemplo, o segundo argumenta que os artefatos técnicos possuem uma certa autonomia para agir e não estão sob o controle dos humanos. Apesar de ambas as perspectivas enfatizarem a interdependência dos humanos e da técnica, Winner (1977) defende que os objetos têm um certo grau de independência, enquanto Latour (1994^b) enfatiza que a ação é resultado de uma “cadeia de mediações”.

Latour (1994^b) considera a tecnologia como um “trabalho congelado”, híbridos que trazem para o presente várias ações passadas de outros actantes, de vários construtores que desaparecem e mesmo assim continuam presentes nos objetos que ajudaram a confeccionar. (Idem, p. 40) Com isso, ele defende que, no lugar de serem considerados meramente “coisas”, os artefatos técnicos merecem ser incluídos em nossa cultura intelectual como completos atores sociais. “Eles mediam as nossas ações? Não, eles são nós.”²⁶ (Idem p. 64)

Ao contrário do que muitos argumentam, Latour (1992) defende que o avanço tecnológico não elimina, e sim mobiliza ainda mais pessoas e, assim, considerar um artefato técnico autônomo seria desconsiderar uma infinidade de humanos que se mobilizaram para criar aquele objeto. (LATOUR, 1992, p. 252) “A ideia bizarra de que a sociedade é feita de relações humanas é uma imagem espelhada de outra ideia não menos bizarra de que as técnicas são constituídas por relações não-humanas.”²⁷ (Idem, p. 239) Para transformar uma ideia em um projeto, um projeto em um experimento piloto, um experimento piloto em um objeto, um grande número de humanos é necessário. (LATOUR, 1994^b, p. 49)

Considere as coisas, e você terá humanos. Considere os humanos, e com este ato você estará muito interessado em coisas. Volte sua atenção para estruturas e objetos duros, e veja eles se tornando gentis, suaves, humanos. Traga então sua atenção para os humanos, e observe eles se tornarem

²⁵ Texto original: “Technical action is a form of delegation that allows us to mobilize, during interactions, moves made elsewhere, earlier, by other actants.” (LATOUR, 1994^b, p. 52)

²⁶ Texto original: “They mediate our actions? No, they are us.” (Idem, p. 64)

²⁷ Texto original: “The bizarre idea that society might be made up of human relations is a mirror image of the other no less bizarre idea that techniques might be made up of nonhuman relations.” (LATOUR, 1992, p. 239)

circuitos elétricos, engrenagens automáticas ou *softwares*.²⁸ (Latour, 2000^a, p. 20)

O autor pontua que a diferença entre um coletivo primitivo e um moderno ou avançado se estabelece exatamente pelo aumento da complexificação das relações entre humanos e não-humanos, sendo que o último é capaz de transladar, cruzar, envolver e mobilizar ainda mais elementos conectados em rede. “Há mais cadeias de ações nos coletivos ‘modernos’, um maior número de não-humanos (máquinas, autômatos, dispositivos) associados uns aos outros, mas não devemos esquecer o tamanho dos mercados, o número de pessoas em suas órbitas, a amplitude da mobilização: mais objetos, sim, mas muitos mais sujeitos também.” (LATOUR, 1994^b, p. 47)

Deste modo, a partir da discussão sobre mediação técnica, devemos conceber o Jornalismo Automatizado como resultado da ação de um híbrido, de uma composição de forças entre actantes, onde um “programa de ação” é estabelecido para a publicação automatizada de notícias em que parte das ações é delegada pelos humanos para os *softwares* NLG. Estas reflexões sobre autonomia tecnológica e mediação técnica/delegação nos instiga, assim, a buscar as redes sociotécnicas que viabilizam a produção de notícias por *softwares* NLG. É importante pontuar que, por esta ser uma inovação recente e em expansão, estamos em um momento privilegiado para observar o fechamento de várias caixas-pretas, em um cenário onde ainda não se verifica a estabilização e a consolidação do uso destes sistemas pelas redações jornalísticas.

2.4 Cartografando as incertezas do Jornalismo Automatizado

Nesta pesquisa, questionamos as afirmações que sustentam que as notícias automatizadas são produzidas a partir de processos algoritmos e sem a intervenção humana, que estaria limitada apenas à etapa de programação dos algoritmos, e assim partimos da premissa de que elas são o resultado de uma complexa associação entre atores humanos e não-humanos. Nesta busca pelas redes sociotécnicas vinculadas a esta inovação, nos inspiramos no STS e optamos por mapear estudos de caso onde efetivamente *softwares* NLG foram implementados em redações jornalísticas para identificar os atores envolvidos no desenvolvimento desta tecnologia, quais declarações emergem no debate em torno do Jornalismo Automatizado, como esta rede se associa a outras visões e concepções de jornalismo. “Estudos de caso são o pão e a manteiga do STS. Quase todos os *insights* no

²⁸ Texto original: Consider things and you will have humans. Consider humans, and you are by that very act interested in things. Bring your attention to bear on hard things, and see them become gentle, soft or human. Turn your attention to humans, and see them become electric, circuits, automatic gears or softwares.” (LATOUR, 2000a, p. 20)

campo surgem deles, e pesquisadores e estudantes ainda se voltam para artigos baseados em casos para compreender ideias centrais e se orientar a partir de problemas.”²⁹ (SISMONDO, 2010, p. viii)

Ao explicar como abordar o STS como uma metodologia, Law (2015) ressalta o papel das práticas sociais e materiais e dos estudos de caso como os fundamentos para o desenvolvimento de uma pesquisa vinculada a este campo, e enfatiza a não divisão entre teoria e empiria. “O STS se desenvolve nas práticas, e é prático em si mesmo.”³⁰ (LAW, 2015, p. 2) Neste cenário, Law (2015) aponta que a Teoria Ator-Rede (TAR) se configura como um método para cartografar como os atores e os objetos são moldados a partir de suas relações.

Seguindo a definição dada pelo o projeto MACOSPOL – *Mapping Controversies in Science & Technology for Politics*, coordenado por Bruno Latour, controvérsias são “cada *bit* de ciência e tecnologia que ainda não está estabilizado, fechado ou ‘encaixapretado’; isso não significa que haja uma disputa feroz, nem que esse conhecimento tenha sido de alguma forma politizado; usamos isso como um termo geral para descrever a incerteza compartilhada.”³¹ Para além da controvérsia em torno da disputa entre jornalistas e robôs, nos apropriamos deste modo de olhar para a tecnociência por reconhecermos que o Jornalismo Automatizado é um fenômeno recente, ainda não estabilizado, e que surge em um contexto repleto de incertezas. Como afirma Latour (2012), “o mundo não se parece com um continente sólido de fatos pontilhados por algumas lagoas de incertezas; é um vasto oceano de incertezas pintalgado de ilhotas de formas calibradas e estabilizadas.” (Idem, p. 348-349)

Neste contexto, do ponto de vista do pesquisador, reconhecer as incertezas de um objeto é de certa forma sair de uma zona de conforto cristalizada nos estudos das ciências sociais em que estamos acostumados a analisar os atores e a taxar-lhes rótulos, como se estivéssemos em uma posição privilegiada e superior que nos dá o “poder” de dizer “verdades” e criar categorias de análises para os acontecimentos. Reconhecer as incertezas nos obriga a observar cuidadosamente os passos destes atores, o que eles dizem e fazem, como eles se associam a outros aliados e que disputas eles travam entre si. Para Venturini (2012), as controvérsias seriam o melhor momento para se observar a construção da vida social e do

²⁹ Texto original: “Case studies are the bread and butter of STS. Almost all insights in the field grow out of them, and researchers and students still turn to articles based on cases to learn central ideas and to puzzle through problems.” (SISMONDO, 2010, p. viii)

³⁰ Texto original: “STS attends to practice, and it is practical itself.” (LAW, 2015, p. 2)

³¹ Texto original: “every bit of science and technology which is not yet stabilized, closed, or ‘black boxed’; it does not mean that there is a fierce dispute, nor that knowledge has been somehow politicized; we use it as a general term to describe shared uncertainty.” Disponível em: <http://www.mappingcontroversies.net:80/Home/PlatformTutorial>

coletivo, quando é possível visualizar o engajamento dos atores na criação e desconstrução de relações, a partir de múltiplas perspectivas e pontos de vista. (VENTURINI, 2012, p. 797)

“Sabemos que não sabemos, mas isso é quase tudo o que sabemos: não há definição melhor para as incertezas.”³² (CALLON, LASCOUMES, BARTHES, 2009, p. 21) Callon, Lascoumes e Barthes (2009) chamam a atenção que, ao contrário do que muitos pensavam há algumas décadas, o desenvolvimento tecnológico e científico não foi acompanhado de um aumento das certezas e, paradoxalmente, ele só gerou mais incertezas. (Idem, p. 18-19) Nestas situações, os autores defendem que a única coisa a se fazer é ampliar os debates e fomentar os questionamentos, em busca de elementos e possíveis respostas. Callon, Lascoumes e Barthes (2009) argumentam que as incertezas são tanto técnicas e científicas quanto sociais, e é muito difícil determinar as fronteiras que delimitam se uma questão é “inquestionavelmente” técnica ou social. “A sociedade pode, de fato, ser tão incerta e imprevisível quanto as entidades não-humanas com as quais escolheu compartilhar o seu destino.”³³ (Idem, p. 26)

No meio de tantas incertezas, a Cartografia de Controvérsias (VENTURINI, 2010, 2012) é uma das principais influências na construção de um olhar metodológico que nos auxilia a observar os atores, a formação de redes, os argumentos e as associações em torno do Jornalismo Automatizado. A Cartografia de Controvérsias é um conjunto de técnicas criadas inicialmente por Bruno Latour como uma versão didática da Teoria Ator-Rede (TAR), com o intuito de auxiliar estudantes e pesquisadores no exercício da observação e descrição de debates sociais em torno principalmente, mas não exclusivamente, de disputas tecnocientíficas. (VENTURINI, 2010, p. 258) Para Latour (2000), acompanhar os períodos iniciais da construção dos fatos tem como vantagem que, no calor da controvérsia, os próprios especialistas podem explicar por que seus oponentes pensam de outro modo. “No estudo de controvérsias e discordâncias, testes de força e translações, sempre parte do pressuposto de que as alegações e contra-alegações são claramente visíveis e estimulam o debate.” (LATOUR, 2000 p. 273)

Venturini (2010, 2012) compara as controvérsias ao magma, e cria um paralelo entre a cartografia destas disputas tecnocientíficas e o movimento das placas tectônicas. Como ele mesmo explica, o magma é uma mistura contendo rochas tanto no estado líquido quanto no estado sólido em constante mutação. “A mesma dinâmica pode ser observada nas controvérsias, a mesma flutuação entre diferentes estados de solidificação. Nesta dinâmica,

³² Texto original: ““We know that we do not know, but that is almost all that we know: there is no better definition of uncertainty.” (CALLON, LASCOUMES, BARTHES, 2009 p. 21)

³³ Texto original: Society may indeed be as uncertain and unpredictable as the nonhuman entities with which it has chosen to share its destiny.” (Idem, p. 26)

o social é incessantemente construído, desconstruído e reconstruído. Isto é o social em ação e é por isso que não temos outra escolha a não ser mergulhar no magma."³⁴ (VENTURINI, 2010, p. 264) Neste paralelo, as cartografias de controvérsias seriam representações semelhantes às das placas tectônicas, onde as colisões representariam os relatos fornecidos por atores discordantes, que se afundam e derretem no calor dos conflitos. "Representar uma controvérsia é como construir em uma falha sísmica. Para suportar a agitação das disputas, as descrições devem ser resistentes a tremores."³⁵ (VENTURINI, 2012, p. 799)

Venturini oferece três recomendações para guiar este processo de observação e descrição dos fatos. A primeira delas é que *o pesquisador não deve restringir a sua observação a nenhuma teoria ou metodologia*, o que não significa que ele esteja livre para não seguir nenhuma teoria ou metodologia pré-determinada. Ao não impor nenhuma filosofia ou procedimento, a cartografia de controvérsias convida os pesquisadores a utilizarem qualquer ferramenta de observação que esteja a seu alcance, misturando várias técnicas sem nenhuma restrição. A orientação é que o cartógrafo se mantenha o mais aberto possível, principalmente no início da investigação, sensível à surpresa e para onde a curiosidade queira o levar. (VENTURINI, 2010, p. 259)

A segunda recomendação é para que *o pesquisador observe o maior número de pontos de vista possível*, sem que isto seja confundido com uma busca pela imparcialidade. É a partir da descrição de um grande número de posicionamentos e perspectivas, sem medo de demonstrar a parcialidade dos atores e de quem observa, que a cartografia se torna cada vez mais diversa. (Idem, p. 260) Venturini (2010) comenta que muitas vezes, Latour e a TAR são acusados de não assumirem um dos lados e de terem um posicionamento político *naïve*, uma visão que não compreende que, para a TAR, o pesquisador não deve buscar a solução de uma disputa coletiva, e nem tentar fechar as discussões em caixas-pretas. O papel de um pesquisador na perspectiva da TAR é de se dedicar a mostrar as várias opiniões e argumentos em jogo, observando e descrevendo todas as parcialidades que encontrar na trajetória de sua pesquisa. Não são os pesquisadores, e sim os atores, os responsáveis por encontrar a solução para uma disputa. (Idem, p. 268)

A terceira recomendação é que *o pesquisador deve ouvir a voz dos atores mais do que suas próprias presunções*, estando atento para sempre avaliar a sua posição com relação aos atores que observa ao longo da cartografia. A precaução é para que os pesquisadores

³⁴ Texto original: "The same dynamic can be observed in controversies, the same fluctuation between different states of solidity. Though this dynamic the social is unremittingly constructed, deconstructed and reconstructed. This is the social in action and that's why we have no other choice than diving in magma." (VENTURINI, 2010, p. 264)

³⁵ Texto original: "Representing a controversy is like building on a seismic fault. To endure the shake of disputes, descriptions must be quake proof." (VENTURINI, 2012, p. 799)

não assumam uma postura arrogante, imprimindo à pesquisa um tom respeitoso que considera o posicionamento dos diversos atores que observa, que podem saber tanto quanto ele sobre os argumentos em disputa em uma controvérsia tecnocientífica. (VENTURINI, 2010, p. 260)

Para orientar a cartografia, Venturini (2010) sugere *cinco lentes de observação* que, como lentes intercambiáveis de um microscópio ou de uma máquina fotográfica profissional, oferecem a possibilidade de olhar para a controvérsia a partir de diferentes ângulos, possibilitando criar descrições mais completas das disputas e dos argumentos em jogo.

1. De declarações para literaturas. Quando se aproxima de uma controvérsia, a primeira impressão é de uma caótica disputa de argumentos e declarações. Neste contexto, a primeira lente de observação na cartografia deve identificar amplamente o universo por onde circula a controvérsia, mapeando uma teia de documentos e referências que revelem os discursos e as ações articuladas nestas literaturas. (Idem, p. 266)

2. De literaturas para atores. A partir da descrição dos argumentos e declarações encontrados nas literaturas, o cartógrafo começa a visualizar os atores envolvidos nesta controvérsia. Assim, a segunda lente de observação é identificar estes atores, que podem ser seres humanos, objetos técnicos, organismos vivos, entidades metafísicas, instituições, qualquer elemento que aja dentro de uma controvérsia (Idem, p. 266)

3. Dos atores às redes. Os atores só podem ser identificados em uma controvérsia porque eles interagem uns com os outros, moldam e são moldados por suas relações e conexões. A cartografia não pode ignorar o dinamismo relacional da controvérsia e, assim, a terceira lente de observação deve olhar para o processo de formação de redes, de conexões sendo estabelecidas e interrompidas, enfatizando o movimento, as mudanças, o fluxo de relações entre os atores. (Idem, p. 266-267)

4. Das redes aos cosmos. A maioria dos atores se associa em torno de grupos – ou ideologias – em busca de algum tipo de estabilidade. Enquanto a vida coletiva é caótica e errática, as ideologias são mais ordenadas, se tornam mais influentes e dão origem a cosmos. A quarta lente de observação da cartografia é associar as ações, os argumentos e as relações

a um significado maior atribuído pelos atores. “Somente vagando entre os cosmos que os cartógrafos sociais podem perceber toda a extensão de suas controvérsias.”³⁶ (Idem, p. 267)

5. Dos cosmos à cosmopolítica. A quinta lente de observação é a mais complicada de todas, pois exige que o cartógrafo abandone a crença de que, por trás das ideologias, existe uma realidade objetiva que não depende do que os atores pensam ou dizem dela. Os atores são levados a crer que vivem em diferentes cosmos e que devem lutar por eles, mas não enxergam que, na realidade, eles vivem em uma constante negociação da existência coletiva. Nesta etapa, o cartógrafo tenta observar como ocorrem as dinâmicas disputas pelo poder entre os cosmos, identificando como a cartografia vai se moldando ao longo do tempo. (Ibidem, p. 267)

É importante deixar claro, como Venturini (2010) chama a atenção, que o objetivo de uma pesquisa que adota a cartografia de controvérsias como metodologia não é a de estabilizar a rede da qual ela se dedica a observar e descrever. A responsabilidade por encontrar soluções para as controvérsias não é do pesquisador, e sim dos atores. O objetivo não é silenciar uma discussão em busca de uma verdade, e sim de mostrar que várias verdades merecem ser observadas e reconhecidas. A qualidade de uma cartografia vai depender da capacidade do pesquisador de acumular dados, documentos, notas, entrevistas, e a sua sensibilidade para monitorar todas estas informações que deixam rastros sobre a riqueza da vida coletiva. (VENTURINI, 2010, p. 269)

2.5 Como cartografamos o Jornalismo Automatizado nesta pesquisa

Esta pesquisa busca a partir de agora, nos próximos dois capítulos, observar os argumentos, os atores, as redes e as associações em torno do Jornalismo Automatizado e, neste contexto, a cartografia de controvérsias nos parece um interessante instrumento de análise por estarmos lidando como um objeto tecnocientífico recente e em plena expansão. Esta inovação atrai a atenção das pessoas de diversas maneiras, tanto positivas quanto negativas, e divide a opinião de jornalistas e leitores que questionam se os computadores são capazes de escrever um texto com uma complexidade próxima à escrita humana. Neste contexto, consideramos a cartografia de controvérsias uma boa ferramenta de análise que

³⁶ Texto original: “Only roaming from cosmos to cosmos, social cartographers can perceive the full extent of their controversies.” (VENTURINI, 2010, p. 267)

nos ajuda a criar descrições mais completas para dar conta da diversidade e complexidade do objeto desta pesquisa.

Inspirada pelas lentes de observação sugeridas por Venturini (2010) como uma forma de criar diferentes modos de olhar que resultem em uma descrição mais completa do fenômeno observado, esta pesquisa realizou dois movimentos diferentes em busca de algumas das redes sociotécnicas vinculadas à esta inovação tecnológica. Podemos dizer que escolhemos dois pontos de entrada distintos para cartografar parte das redes sociotécnicas vinculadas a esta inovação. Num primeiro momento, nosso olhar se voltou para um cenário mais acadêmico que descreve pesquisas e experimentos que, de alguma forma, se articulam com o Jornalismo Automatizado, o que defendemos ser um campo de estudos em formação. Num segundo momento, nosso olhar se dirige para a descrição de três experiências pioneiras de produção automatizada de notícias, todas elas localizadas nos Estados Unidos, que nos ajudam a visualizar como empresas de mídia, *startups*, universidades e outras instituições de pesquisa que fornecem dados estatísticos se articulam em torno desta inovação.

Cada um destes movimentos demonstra um esforço de abertura de diferentes caixas-pretas associadas à produção das notícias automatizadas na busca pelas redes sociotécnicas que deem visibilidade à dimensão humana por trás desta inovação tecnológica. Estes movimentos também acabam por se associar a diferentes concepções de jornalismo vinculadas à implementação de *softwares* NLG nas redações, com relações intrínsecas e imbricadas com as Ciências da Computação, o Jornalismo Computacional e o Jornalismo de Dados. Como corpus de pesquisa, a cartografia foi realizada com base em uma seleção de 54 notícias, reportagens e *posts* em blogs, cinco vídeos no *YouTube*, doze artigos científicos, uma *API* e duas patentes de *softwares* NLG, sendo que todos estes documentos foram publicados entre 2009 e 2017.

Assim, num primeiro movimento, buscamos nas pesquisas tanto do jornalismo quanto das Ciências da Computação um embasamento científico para esta investigação, onde reconhecemos que o Jornalismo Automatizado é ainda um campo de estudos em formação, mas que está diretamente ligado a pelo menos a outras duas áreas: o Jornalismo Computacional e o *Natural Language Generation*, um subcampo da Inteligência Artificial e da Linguística Computacional. Identificamos, nesta cartografia, algumas publicações científicas que, desde 2012, tentam compreender como a produção automatizada de notícias foi adotada por redações, instigadas principalmente em compreender como esta inovação é recebida tanto por jornalistas quanto por leitores. É importante pontuar que o jornalismo já convive com a automação há pelo menos quatro décadas, sendo que a primeira aplicação deste tipo de tecnologia foi para a redação de previsões do tempo. No capítulo 3, descrevemos ainda o funcionamento dos *softwares* que produzem as notícias automatizadas, em uma cartografia

realizada a partir de publicações acadêmicas, do registro de duas patentes pela *startup Narrative Science* e da versão de demonstração do *software Wordsmith*, desenvolvido pela *startup Automated Insights*.

E num segundo movimento, o foco desta pesquisa foi na cartografia de três estudos de caso onde veículos jornalísticos adotaram recentemente *softwares* para a produção automatizada de notícias para a cobertura de esportes, finanças, educação, homicídios e terremotos, todos eles localizados nos Estados Unidos. O primeiro estudo de caso descreve a trajetória da *startup Narrative Science* desde o protótipo *StatsMonkey* que deu origem à empresa, desenvolvido em uma disciplina de pós-graduação na *Northwestern University*, passando pela polêmica afirmação de que um “robô” em breve poderia ganhar um *Pulitzer*, atravessando as parcerias com a *Forbes* e com a *ProPublica* até o anúncio, em 2016, de que eles não atendiam mais nenhum cliente do ramo jornalístico. O segundo estudo de caso gira em torno de outra *startup* que surgiu em 2010 como *StatSheet*, mas que posteriormente mudou seu nome para *Automated Insights* e estabeleceu uma parceria com a agência *Associated Press* para a redação automatizada de notícias sobre os resultados financeiros de empresas listadas na bolsa. O terceiro e último estudo de caso narra como o jornal *Los Angeles Times* adotou o Jornalismo Automatizado em dois projetos desenvolvidos internamente pela equipe de jornalistas de dados, um para a cobertura de homicídios denominado *The Homicide Report*, e outro para a cobertura de terremotos denominado *Quakebot*.

Ao selecionamos as reportagens, *posts* de *blogs* e vídeos para mapearmos os atores, as declarações e as redes em torno do Jornalismo Automatizado, temos consciência de que nossa cartografia é limitada às escolhas tomadas ao longo deste processo, e que estamos cartografando apenas parte destas associações. As descrições que surgem do relato dos estudos de caso nesta pesquisa produzem um retrato subjetivo que reflete o modo como olhamos para este objeto e, neste sentido, não temos a pretensão e nem a arrogância de considerar este o único modo, ou o melhor modo, ou a forma mais completa de analisar as notícias automatizadas. Quando optamos pelo STS, pela Teoria Ator-Rede e pela Cartografia de Controvérsias, estamos cientes de que estamos fazendo escolhas e definindo um modo de olhar, um enquadramento metodológico para analisar as notícias escritas por “robôs”.

Estas escolhas foram motivadas, principalmente, por uma vontade de observarmos o Jornalismo Automatizado de maneira leve, sem um percurso pré-definido que reafirmasse, sob uma perspectiva antropocêntrica, convicções pré-determinadas de superioridade da mente humana tanto na escrita quanto na redação de notícias. Com estas escolhas, nossa intenção foi de buscar um ponto de vista o mais simétrico possível que tivesse ao menos a pretensão de olhar para o Jornalismo Automatizado sem pré-conceitos, e conseguisse

observar como estas notícias são construídas a partir de redes heterogêneas e de um “programa de ação” engendrado por humanos – jornalistas, empreendedores, programadores – e não-humanos – algoritmos, *softwares*, bancos de dados computadores.

3 Jornalismo Automatizado: cartografando tecnologias e pesquisas

Nesta primeira etapa da cartografia, buscamos situar o Jornalismo Automatizado em um contexto acadêmico e científico, onde sugerimos que este seja um campo de estudos em formação que dialoga com as Ciências da Computação e com o Jornalismo Computacional, com uma forte articulação com universidades, *startups*, patentes e empresas de mídia.

Este capítulo é dividido em quatro partes, e a primeira delas relata como a automação já está presente no jornalismo há algumas décadas. Desde os anos 1970, surgiram os primeiros *softwares* aptos a escrever previsões do tempo automatizadas, que foram a primeira aplicação desta tecnologia no jornalismo. Estes sistemas estavam vinculados a pesquisas acadêmicas ligadas a um subcampo da Inteligência Artificial denominado *Natural Language Generation* (NLG). Na segunda parte, dirigimos o nosso olhar para as materialidades e descrevemos, a partir da API do *Wordsmith* e de alguns registros de patente da *Narrative Science*, como os *softwares* NLG funcionam, suas articulações com o banco de dados e com o texto formulaico que estrutura as narrativas automatizadas.

Apesar de ainda ser muito recente, o Jornalismo Automatizado tem atraído cada vez mais a atenção de pesquisadores no mundo inteiro, que buscam através de pesquisas e experimentos compreender como esta tecnologia se insere nas redações. Na terceira parte, discutimos alguns dos resultados destas pesquisas que focam, principalmente, na reação de jornalistas e leitores às notícias automatizadas. Por fim, a quarta parte relata os esforços acadêmicos para encontrar uma definição para o Jornalismo Automatizado e como estas pesquisas dialogam frequentemente com o Jornalismo Computacional e com as discussões em torno do crescente uso de algoritmos e *big data* pelas redações.

3.1 Décadas de automação no jornalismo

Muitas vezes ainda tomados como algo que está por vir, a produção automatizada de textos jornalísticos já é uma realidade há alguns anos. Desde as décadas de 1950 e 1960, um campo de estudos das Ciências da Computação denominado *Natural Language Generation* (NLG) se dedica a realizar experiências de tradução e criação de pequenas frases com estruturas gramaticais simples, sistemas que se tornaram mais complexos a partir da década de 1970 quando apareceram *softwares* capazes de escrever textos. (REITER, DALE, 2000, p. 19) Foi nesta época que surgiram os precursores das primeiras aplicações jornalísticas deste tipo de tecnologia: as previsões do tempo escritas de forma automatizada por um computador, como o experimento desenvolvido por Harry R. Glahn nos Estados Unidos que

utilizava como *input* o banco de dados com informações meteorológicas da *ESSA Weather Bureau's National Meteorological Center* (GLAHN, 1970, p. 1126).

Como é possível observar na figura 3, as primeiras previsões do tempo escritas por computador consistem em textos simples, que seguem uma mesma estrutura linguística alternando algumas palavras, como o nome da cidade, a temperatura, a probabilidade ou não de chuvas e de neve. Neste experimento, Glahn (1970) conclui que os computadores estão aptos a produzir previsões do tempo de forma automatizada, e que os textos têm qualidade suficiente para serem aceitos pelo público. Em tom sugestivo, ele propõe que o sistema desenvolvido em sua pesquisa comece a gerar previsões do tempo automáticas para todas as cidades dos Estados Unidos, mas sem tirar a autonomia dos profissionais que atuam nos centros meteorológicos locais de modificarem e adaptarem os textos como desejarem. (Idem, p. 1128-1129)

Figura 3 - Exemplo das primeiras previsões do tempo escritas por um computador

MAR 17, 1970	TDL EXPERIMENTAL FORECASTS
GOOD MORNING. THE TECHNIQUES DEVELOPMENT LABORATORY BRINGS YOU THE LATEST FORECAST FOR WASHINGTON, D. C. AND VICINITY. MOSTLY SUNNY THIS MORNING WITH A FEW MORE CLOUDS THIS AFTERNOON. SOMEWHAT WARMER TODAY, MAXIMUM TEMPERATURE 47 DEGREES. NORTHWESTERLY WINDS OF 5 MPH THIS MORNING BECOMING LIGHT AND VARIABLE BY AFTERNOON. ONLY 2 PERCENT PROBABILITY OF PRECIPITATION TODAY.	
GOOD MORNING. THE TECHNIQUES DEVELOPMENT LABORATORY BRINGS YOU THE LATEST FORECAST FOR ATLANTA AND VICINITY. PARTLY CLOUDY THIS MORNING BECOMING CLOUDY THIS AFTERNOON. LITTLE CHANGE IN TEMPERATURE TODAY, HIGH OF 53 DEGREES. SOUTHEASTERLY WINDS 15 MPH, WITH 15 PERCENT PROBABILITY OF RAIN AND 2 PERCENT PROBABILITY OF SNOW.	

Fonte: GLAHN, 1970, p. 1128

Na década de 1990, um dos desafios era desenvolver sistemas multilíngues como, por exemplo, o *FoG*, um *software* desenvolvido em 1993 pela *CoGenTex* para a empresa pública que fornece previsões meteorológicas no Canadá que, apesar de produzir textos simples, se destacou por escrever tanto em inglês quanto em francês. (GOLDERG et al, 1994).

Um outro exemplo semelhante ao *FoG* é o *SumTime-Mousam*, um software desenvolvido por professores da *University of Aberdeen*³⁷, na Escócia, para a empresa *Weathernews*, que fornece imagens e informações meteorológicas para canais de TV, estações de rádio e *sítes* de vários países. (SRIPADA, REITER, DAVY, 2003, p.4) Ao avaliarem os textos produzidos por este sistema, em uma análise comparada com textos escritos por humanos, pesquisadores observaram que os leitores muitas vezes preferiam as previsões do tempo escritas pelo *SumTime-Mousam* por conta de uma melhor escolha de palavras para comunicar de forma precisa números e dados estatísticos. (REITER et al, 2005, p.163) Semelhante à experiência de Glahn (1970), esta pesquisa demonstrou que *softwares* NLG estão aptos a gerar textos sobre previsões do tempo com qualidade.

No Brasil, Nilson Lage (1997, 2003) foi um dos pioneiros nesta discussão, ao desenvolver uma proposta de automação da redação de *leads*³⁸ a partir de regras linguísticas de sintaxe e semântica em que demonstra o caráter repetitivo e previsível desta estrutura inicial da narrativa jornalística. Para Lage (1997), um dos aspectos fundamentais para a aplicação de Inteligência Artificial ao jornalismo é a distinção entre notícia e reportagem, sendo a notícia uma narrativa mais simples, menor e fundamentada no relato de um fato ou acontecimento, e a reportagem uma narrativa mais aprofundada dedicada à discussão sobre um assunto. Quinze anos antes do anúncio da parceria entre a *Forbes* e a *startup Narrative Science*, fato que gerou grande repercussão mundial em torno da redação de notícias por “robôs”, Lage (1997) já defendia que a automação do discurso jornalístico só seria aplicável para a redação de notícias – e não de reportagens – sobre assuntos como “observações meteorológicas(...), produção de anúncios classificados; avisos de tráfego aéreo, terrestre ou marítimo; noticiário sobre a variação das cotações em bolsas de ações, moedas ou mercadorias; alterações nas colocações em *rankings* ou campeonatos”. (LAGE, 1997, p. 1)

Estes exemplos desmistificam a ideia de que o Jornalismo Automatizado seja uma novidade ou algo que se materializará apenas no futuro, como muitos acreditam. Como defende Linden (2016), a automação de tarefas jornalísticas já acontece há pelo menos quatro décadas e esta recente inovação talvez seja “apenas mais um passo na história do jornalismo-maquínico”. (LINDEN, 2016, p. 1) Por outro lado, apesar de já estar sendo desenvolvido há

³⁷ Em 2009, estes mesmos professores vinculados à *University of Aberdeen* fundaram uma *startup* - a *Data2Text* – com o objetivo de fornecer previsões do tempo automatizadas, e um dos clientes que contratou este serviço foi o *MeteoGroup*, no Reino Unido³⁷. Em maio de 2012, a *Data2Text* mudou seu nome para *Arria NLG* e, além das previsões do tempo, passou a fornecer textos automatizados para outros setores não-jornalísticos, como relatórios de monitoramento de UTI Neonatal.

³⁸ Primeiro parágrafo de uma notícia que se destaca por sua estrutura lógica que informa, seguindo a síntese de Laswell, quem fez o que, a quem, quando, onde, como, por que e para quê. De acordo com Nilson Lage, o “*Lead* é a abertura de uma notícia: proposição completa, constituída de sujeito, verbo, complementos e circunstâncias, que se inicia pela notação mais importante ou interessante e que pode apresentar-se, no nível de sua realização sintática, por um ou mais períodos no mesmo parágrafo lógico.” (LAGE, 1997)

vários anos, é inegável o fato de que o Jornalismo Automatizado ganhou maior destaque a partir de 2010, quando as *startups Narrative Science* e *Statsheet* começaram a produzir de forma massiva notícias sobre jogos esportivos universitários nos Estados Unidos.

3.2 Como funcionam os *softwares de Natural Language Generation*

Voltando o nosso olhar para as APIs (*Application Programming Interface*), os *softwares de Natural Language Generation* (NLG) fazem parte de um “subcampo da Inteligência Artificial e da Linguística Computacional que foca no desenvolvimento de *softwares* para produzir textos em linguagem natural”³⁹ (REITER, DALE, 2000 p. 1). Estes sistemas produzem narrativas através do processamento de outras fontes de informação, como tabelas e bancos de dados, e estão aptos a produzir de forma automática documentos, relatórios, notícias, desde que estes sigam uma estrutura repetitiva e pré-determinada. De forma simplificada, podemos dizer que *softwares* NLG funcionam como um tradutor, convertendo dados em textos.

No jornalismo, estes *softwares* recentemente vêm sendo implementados para a produção de notícias cuja estrutura é baseada em informações estatísticas que são fornecidas por bancos de dados, como nos casos descritos a seguir nesta dissertação que abrangem coberturas de jogos esportivos, resultados financeiros de empresas listadas na bolsa, crimes e terremotos. Além de notícias, *softwares* NLG também são utilizados para a produção de relatórios de *Business Intelligence*, para a descrição de produtos em sites de *e-commerce*, para a descrição de resultados de exames médicos em um tom amigável para o paciente, entre outras aplicações. (REITER, DALE, 1997 p. 2-3)

Apesar de cada sistema ter sua especificidade, uma explicação simplificada sobre seu funcionamento daria conta de que todos eles possuem um texto padrão com algumas lacunas em branco, que serão preenchidas com informações fornecidas pelo banco de dados. A partir da atualização deste banco de dados, os algoritmos analisam as informações e preenchem as lacunas em branco, exportando uma narrativa. Em seu artigo “*Mapping the field of Algorithmic Journalism*”, Dörr (2015) explica quais seriam as etapas para um *software* NLG receber os dados (*input*), processá-los e publicá-los no formato de uma notícia (*output*).

começando pelo *input* de dados (*request*), algoritmos aplicam operações estatísticas para selecionar informações de um banco de dados e avaliá-los em termos de relevância. O sistema NLG deve então decidir quais estruturas linguísticas (palavras, sintaxe, sentenças) serão utilizadas para comunicar a informação desejada. (...) Como observado por Reiter (2010) o processo de

³⁹ Texto original: “a subfield of artificial intelligence and computational linguistics that focused on computer systems that can produce understandable texts in English or other human languages.” (REITER, DALE, 2000)

geração de conteúdo é complexo e requer muitas decisões, incluindo escolhas lexicais (de qual conteúdo e palavras devem ser usadas para expressar os conceitos principais referentes aos dados) (...) O resultado (*output*) é um texto em linguagem natural. Após este processo de geração, os textos são publicados automaticamente em veículos de comunicação *online* ou *offline*.⁴⁰ (DÖRR, 2015, p. 6)

No livro *Building Natural Language Generation Systems*, Reiter e Dale (2000) dão explicações mais aprofundadas sobre como se dividiria a arquitetura de um *software* NLG, e destacam que o processamento de linguagem natural destes sistemas pode ser dividido em três etapas:

1. **Planejamento do documento:** na primeira etapa, é determinado o conteúdo, quais dados serão utilizados e a estrutura do texto;
2. **Micro-planejamento:** na segunda etapa, o sistema decide quais palavras, estrutura sintática, lexical, como as frases serão combinadas e como as referências serão usadas para comunicar o conteúdo e a estrutura escolhidas na primeira etapa;
3. **Finalização:** na terceira etapa, o *software* mapeia todas as representações abstratas geradas no micro-planejamento e processa as informações no formato de um texto. (REITER, DALE, 2000 p. 59)

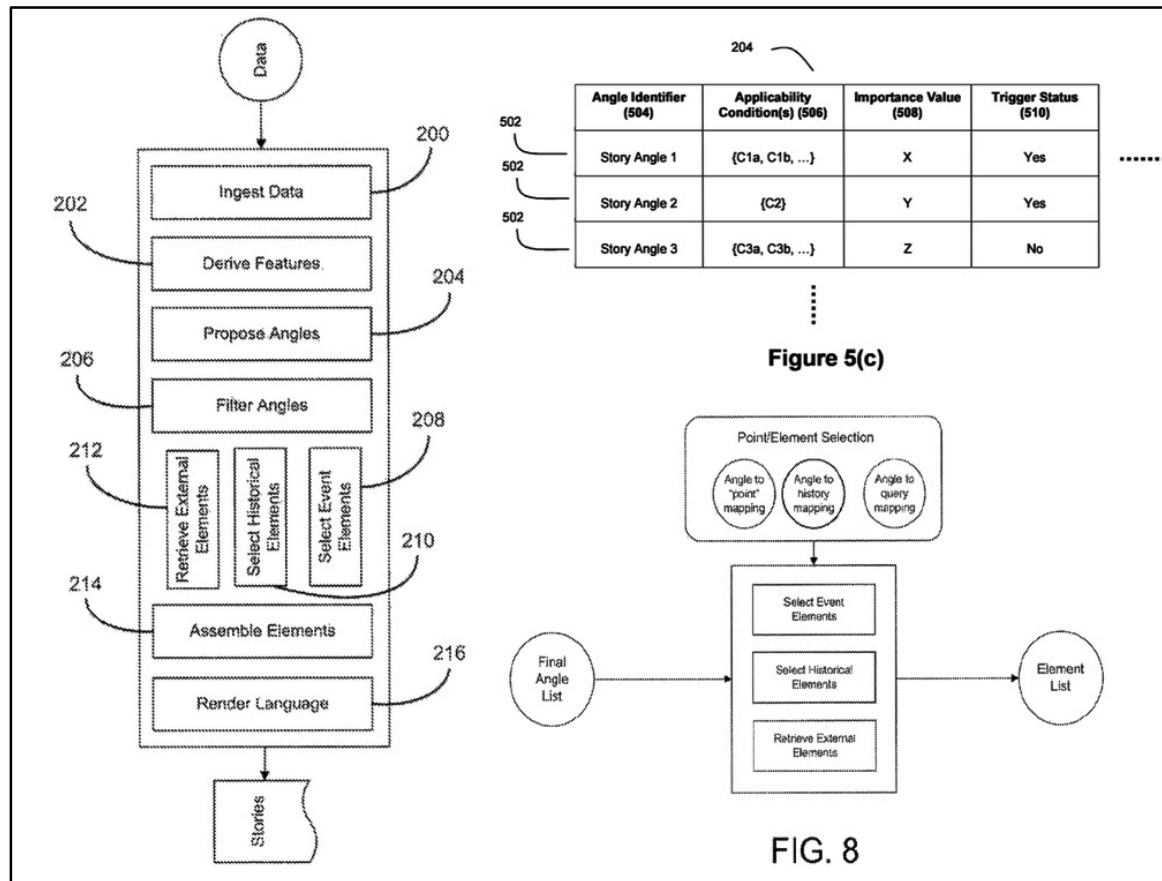
Cada uma dessas etapas possui ainda sub-etapas que complexificam as estruturas dos *softwares* NLG em um esforço de produzir textos mais elaborados e personalizados que não reproduzam uma mesma estrutura gramatical e apenas modifiquem os dados estatísticos. Nestas sub-etapas, são adicionadas outras variáveis, como ângulos de abordagem, elementos históricos, sinônimos de palavras e expressões, criando narrativas cada vez mais complexas. É possível observar estes processos na análise dos fluxogramas obtidos em dois registros de patente da empresa *Narrative Science* (2014), responsável pelo desenvolvimento do *software Quill*. (Figura 4)

Ao observar o primeiro fluxograma à esquerda da Figura 4, é possível compreender de forma mais detalhada a lógica de funcionamento dos *softwares* NLG, que começa com a inserção de dados, a identificação de características destas informações, a análise de quais ângulos se aplicam a estes dados (204), filtragem desses ângulos (206), selecionando os que se aplicam e aqueles que serão descartados, a inclusão de elementos externos (212),

⁴⁰ Texto original: "Starting from the input level (request), algorithms apply statistical operations to select elements from a basic data-set and assign relevance to them. The NLG system must decide which linguistic structures (words, syntax, sentences) should be used to communicate the desired information. (...) As Reiter (2010) notes, this content generation process is complex and requires many decisions, including the lexical choice (choosing which content and words should be used to express domain concepts and data reference). (...) The result (output) is a text in natural language. After this generation process, the texts are published mainly automatically on online or offline news outlets." (DÖRR, 2015, p. 6)

históricos e relacionados àquele evento, a organização destes elementos, a renderização destas informações para então se gerar a narrativa.

Figura 4 – Fluxogramas retirados de dois registros de patentes da *Narrative Science*



Fonte: NARRATIVE SCIENCE INC., 2014

Os outros dois fluxogramas da Figura 4 detalham duas etapas específicas do fluxograma à esquerda. No fluxograma mais acima à direita, identificado como 5(c), é possível observar a lógica de como são definidos os ângulos de abordagem da notícia automatizada, com a criação de operações condicionais que definem qual padrão deve ser aplicado ao texto, a definição de grau de importância de cada ângulo, e se aquela condicional está presente ou não em um determinado processamento de dados, em uma operação algorítmica simples determinada por Sim/Não. O último fluxograma abaixo à direita (FIG. 8) esquematiza como é realizada a inclusão de elementos históricos, externos e do evento que serão definidos na etapa de micro-planejamento, para então serem aplicados na narrativa.

Em novembro de 2015, no início desta dissertação, tivemos acesso a uma versão “demo” do *software Wordsmith*, produzido pela *Automated Insights* e utilizado pela *Associated Press*, que nos possibilitou observar como as notícias automatizadas são produzidas a partir de um *template* e de dados estruturados em tabelas CSV. Na seção “ajuda”, o *software*

apresenta o “*Wordsmith Project Roadmap*”⁴¹, onde descreve os processos de customização do *software* para a produção de notícias automatizadas, em um roteiro dividido em dez etapas:

Planejamento da narrativa automatizada

1. Criar um *template*: a primeira etapa é pensar no que você quer criar com o *Wordsmith* e por quê. Algumas perguntas que você precisa responder nesta etapa são: O que você quer criar? O que é importante incluir nesta narrativa? Qual o público-alvo do texto? Onde este texto será publicado? Em qual formato? Qual a frequência que você irá utilizar o *Wordsmith*? Quantas narrativas você quer produzir de uma só vez?⁴²

2. *Brainstorm*: a segunda etapa é anotar as ideias que você tem de *template*, sempre pensando na sua audiência e no tipo de narrativa que você deseja produzir a partir de quais tipos de dados.

3. Esboçar a lógica: na terceira etapa, é necessário pensar na estrutura do texto, onde podem ser adicionadas “lógicas condicionais” que serão aplicadas aos dados. Isto permitirá que um mesmo *template* gere narrativas diferentes dependendo dos cenários específicos identificados a partir da análise e do processamento de dados. Estas “lógicas condicionais” serão posteriormente transformadas em regras (*branches*) do *template*.

Figura 5 - Tabela CSV do *Wordsmith* com resultados financeiros de fundos de investimento

qtr	year	qtr_end_date	fund_name	fund_quarter_return	benchmark_name	benchmark_qtr_return	previous_quarter	previous_quarter_year	previous_quarter_return
1	2016	31-Mar	Fund 1	1.24	Benchmark 1	6.36	4	2015	6.2
4	2013	31-Dec	Dornis Fund	4.14	Benchmark B	0.97	3	2013	4.3
3	2016	31-Dec	Treeselectrics Index	-4.32	Benchmark C	1.35	3	2016	1.13
3	2015	30-Sep	Isholding Fund	6.36	Benchmark D	1.24	2	2015	5.45
3	2015	30-Sep	Fasela Fund	-8.4	Benchmark E	-7.3	2	2015	-7.33
3	2015	30-Sep	Zumtex Index	2.25	Benchmark F	3.25	2	2015	0.98
4	2015	31-Dec	Villaflex Fund	2.53	Benchmark H	2	3	2015	-0.13
3	2015	30-Sep	Acecane Index	-4.73	Benchmark I	-3.78	2	2015	0.12
3	2015	30-Sep	Inchdom Index	2.23	Benchmark J	2.02	2	2015	1
1	2015	31-Mar	Dingholding Fund	-5.11	Benchmark B	-6.44	4	2014	-3.68

Fonte: Reprodução do site da *Automated Insights*⁴³

⁴¹ *Wordsmith Project Roadmap*. Disponível em: <https://wordsmithhelp.readme.io/docs/wordsmith-project-roadmap>. Acesso em 22/05/2017.

⁴² *Designing a Template*. Disponível em: <https://wordsmithhelp.readme.io/docs/designing-a-template-for-your-audience>. Acesso em 22/05/2017.

⁴³ <https://wordsmith.automatedinsights.com/gallery/quarterly-fund-report> Acesso em: 22/05/2017

Dados

4. Dados estruturados: a quarta etapa consiste na definição da estrutura dos dados necessários para gerar as narrativas. Esta etapa deve exigir um pouco de análise e pode, por exemplo, incluir cálculos e/ou agregar dados.

A Figura 5 mostra parte de uma tabela CSV utilizada pelo *Wordsmith* para gerar relatórios trimestrais de resultados financeiros de fundos de investimento, um dos exemplos disponíveis no site da *Automated Insights* e que se aproxima muito do contexto em que este *software* é utilizado pela *Associated Press*.

Escreva seu *Template*

5. Projeto: a quinta etapa é criar um novo projeto no *Wordsmith* e um novo *template* vinculado aos dados e à tabela estruturada na quarta etapa.

6. Branches: na sexta etapa, a partir das “lógicas condicionais” que foram esboçadas na terceira etapa, é necessário definir as regras (*branches*) que vão alterar algumas estruturas de seu *template* a partir de diferentes cenários identificados no processamento de dados. Se usar uma mesma lógica mais de uma vez, use também a fórmula do Sim/Não para identificar diferentes padrões em seus dados e aplicá-los a suas narrativas.

7. Formatar: a sétima etapa consiste em inserir os dados no *template* e formatar como eles deverão aparecer nas narrativas. Nesta etapa, é possível ainda criar fórmulas para qualquer cálculo necessário para o processamento adequado dos dados.

8. Variáveis: na oitava etapa, é possível adicionar sinônimos para palavras e expressões e novas regras condicionais com o intuito de criar mais variações nas narrativas.

Narrativas

9. Revisar os textos: a nona etapa consiste na revisão das narrativas na página de *preview* do site, para verificar possíveis erros gramaticais e se as regras (*branches*) estão modificando os textos da maneira prevista.

Publicação

10. Publicar: na décima e última etapa, é hora de decidir como serão exportadas as narrativas. É possível publicar os textos diretamente da plataforma do *Wordsmith*, utilizando uma das integrações oferecidas pelo *software*, ou então fazer o *download* manual dos textos para uma posterior publicação em outras aplicações ou plataformas.

Figura 6 - Exemplo de *template* do Wordsmith para gerar relatórios sobre resultados financeiros de fundos de investimento a partir dos dados da tabela da Figura 5

233 BRANCHES	324 SYNONYMS	469 DATA VARIABLES
------------------------	------------------------	------------------------------

Insert Data
Add Synonym
Add Branch
More ▾

With a return of 1.24 percent in the 1 quarter, the Fund 1 continued to perform above its benchmark (Benchmark 1, +6.36 percent) for the 1 consecutive quarter. While the fund did not beat its benchmark like it did last quarter, it did improve overall from last quarter's 6.2 percent return, marking the 1 consecutive quarter the fund increased its return.

Strong stock selection drove 1 quarter fund performance, with 3 of the fund's sectors posting returns that surpassed their index counterparts. Sector performance was very strong, with all 5 sectors contributing positive returns. Top performers included: industrials (+5.2 percent), consumer discretionary (+2.7 percent), materials (+7.3 percent).

The fund's only positive sector during the quarter was industrials (+5.2 percent), consumer discretionary (+2.7 percent), materials (+7.3 percent). Performance was overwhelmingly poor this quarter, with all 5 sectors contributing negative returns to the fund. Among the worst detractors were financials (-0.3 percent), health care (-1.7 percent), information technology (-1.0 percent). They combined to detract -1 percentage points from the fund's return.

For the year, the Fund 1 has generated a return of 3.97 percent, with the Benchmark 1 also reporting a return of 1.75 percent. The fund's performance was significantly better than last year at the end of Q1, when it returned 1.5 percent. utilities (+8.3 percent) were the top contributors to the fund's return, while information technology (-6.3 percent), consumer discretionary (-5.4 percent), health care (-4.7 percent) were the top detractors over the 12-month period.

Fonte: Reprodução do site da *Automated Insights*⁴⁴

A Figura 6 nos mostra um exemplo de *template* disponível no site do *Wordsmith* para a redação de relatórios sobre resultados financeiros de fundos de investimento, que utiliza como banco de dados uma tabela CSV como a da Figura 5. Neste *template*, podemos observar que existem 233 regras (*branches*), 324 possibilidades de sinônimos e 469 variáveis para um texto final que terá apenas quatro parágrafos. Ao lermos o texto do *template*, vemos as indicações dos locais exatos onde os dados estruturados na tabela serão inseridos, como por exemplo onde está escrito “Fund 1” será o nome do fundo de investimentos em destaque, “Benchmark 1” também será um dado vinculado à tabela, e assim sucessivamente.

3.3 Softwares NLG no jornalismo: experimentos e legitimação

Desde 2012, foram publicadas várias pesquisas sobre a aplicação de *softwares* NLG no contexto jornalístico, em um campo impulsionado principalmente por pesquisadores norte-americanos e europeus. Os principais questionamentos desta literatura científica dizem

⁴⁴ <https://wordsmith.automatedinsights.com/gallery/quarterly-fund-report> Acesso em: 22/05/2017

respeito a quais características um texto escrito por um “robô” possui, quais os impactos desta inovação tecnológica nas redações, qual a opinião dos jornalistas a respeito das notícias automatizadas, como distinguir se uma notícia foi escrita por um repórter ou por um *software*.

Com o lançamento do *Statsheet* em novembro de 2010, uma rede com 345 sites para a cobertura dos jogos de basquete universitário dos Estados Unidos que estimava a publicação de 15 mil notícias por mês sem a intervenção humana, as primeiras reações dos jornalistas dão conta de que o “repórter-robô” é um novo competidor que coloca em risco a manutenção do emprego principalmente de profissionais que desempenham tarefas repetitivas e rotineiras. (VAN DALEN, 2012) Neste cenário, Van Dalen (2012) aponta que a produção automatizada de notícias desafia os jornalistas a repensarem seu papel e suas habilidades profissionais. Alguns destes jornalistas, como cita o autor, consideravam inclusive esta “uma oportunidade para tornar o jornalismo mais humano”⁴⁵ (Idem, p. 648) pois, para diferenciarem o seu trabalho de um texto escrito por um algoritmo, eles precisariam explorar mais a criatividade, recursos como o humor e a ironia, e oferecerem coberturas mais aprofundadas.

De um modo geral, os estudos identificados em nossa cartografia concluem que os jornalistas consideram as notícias automatizadas chatas e cheias de clichês, mas concordam que esta tecnologia tem o potencial para modificar a forma como eles trabalham. Se por um lado os mais pessimistas temem perder o emprego e preveem uma crescente personalização e polarização dos conteúdos, por outro os mais otimistas visualizam um futuro onde a automação liberará os repórteres de tarefas mecânicas e rotineiras para se dedicarem mais a reportagens investigativas, o que poderia inclusive impactar positivamente na qualidade das matérias jornalísticas. (VAN DALEN, 2012, p. 653; CARLSON, 2014) Estes estudos apontam ainda a possibilidade destes sistemas se tornarem aliados dos jornalistas, ao criarem alertas que identificam detalhes ou padrões nas informações contidas em uma planilha, por exemplo, que poderiam passar despercebidas pelo olhar do jornalista. (CARLSON, 2014, p. 14)

Para Van Dalen (2012), a geração de conteúdo automatizado é uma tendência que precisa ser levada a sério, pois seus baixos custos de produção e a rapidez com que esta tecnologia consegue publicar um grande volume de notícias pode ser uma oportunidade para as empresas jornalísticas ampliarem suas margens de lucro, explorando a “cauda longa” (ANDERSON, 2006) com a cobertura de novas temáticas, como jogos esportivos universitários.

⁴⁵ Texto original: “Journalists see ‘robot journalism’ as an opportunity to make journalism more human.” (VAN DALEN, 2012, p. 648)

Uma outra preocupação levantada pelos pesquisadores é o quanto o Jornalismo Automatizado contribuiria para o processo de comoditização das notícias (CARLSON, 2014), fenômeno já identificado nas redações *online* quando Boczkowski (2010) chama a atenção para o crescente grau de homogeneização e imitação dos conteúdos publicados, mas que ganha uma outra dimensão quando são introduzidos *softwares* aptos a publicar milhares de notícias seguindo *templates* formulaicas vinculadas a bancos de dados estruturados. Para Carlson (2014), o criador de conteúdo automatizado está na interseção entre o jornalista e o analista de *big data*, onde há um estímulo a serem formadas parcerias com profissionais especializados em programação.

Em uma outra pesquisa, Thurman et al (2017) identificam que alguns jornalistas se preocupam com a falta de uma abordagem humana para os textos orientados por dados e, mesmo que o Jornalismo Automatizado tenha um custo reduzido, há um consenso de que sua aplicação é limitada e muitos profissionais não acreditam que esta tecnologia tenha potencial para modificar a dinâmica econômica da indústria. (THURMAN et al., 2017, p. 10) Linden (2016) aponta ainda que é preciso reconhecer que a redução dos postos de trabalhos nas redações em todo o mundo é um reflexo das mudanças no comportamento dos leitores e do modelo de negócios das empresas de mídia que não foram capazes de se adaptar de forma sustentável ao processo de digitalização, onde a culpa, pelo menos até agora, não pode recair na automação.

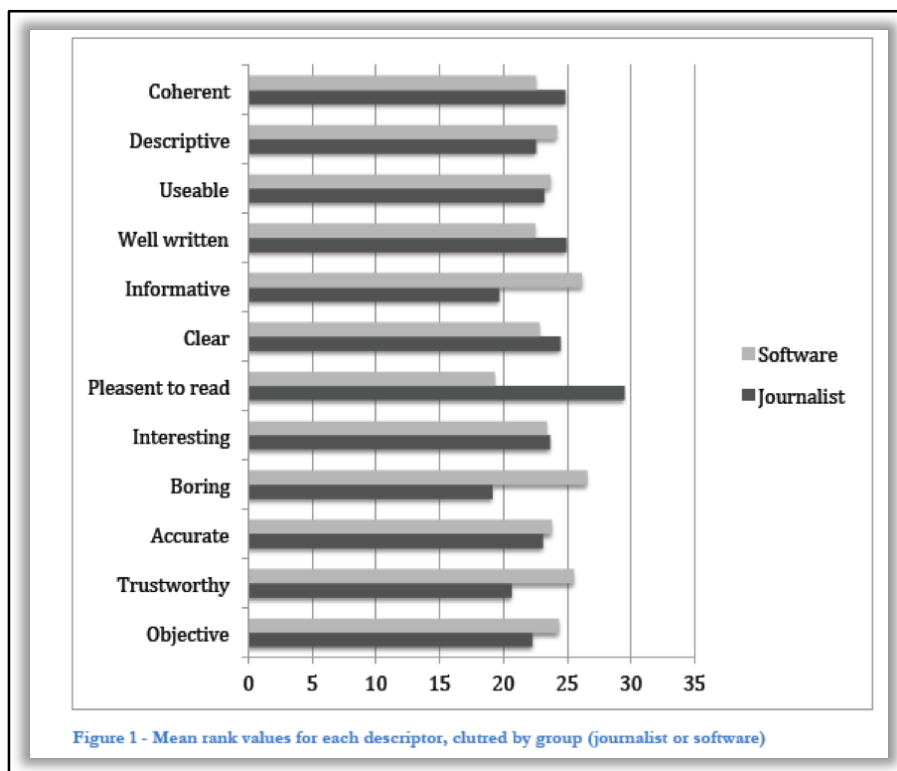
Do ponto de vista dos leitores, pelo menos quatro pesquisas identificadas na cartografia avaliaram a percepção das notícias automatizadas em uma análise comparada com notícias escritas por um jornalista, e todas elas seguiram como referência os critérios definidos por Sundar (1999). A primeira delas (CLERWALL, 2014) foi realizada com 46 estudantes de graduação que analisaram duas notícias esportivas seguindo doze indicadores sem saberem previamente se o texto fora escrito por um repórter ou por um *software*.

Como aponta a Figura 7, os resultados desta pesquisa sugerem que a notícia automatizada é percebida como descritiva, informativa, chata, confiável e objetiva, enquanto a notícia escrita por um jornalista é melhor avaliada nos quesitos coerência, clareza, bem escrita e prazerosa de ler. Apesar deste ser um estudo limitado a partir da avaliação de poucas pessoas, o que mais chama a atenção nesta tabela é a proximidade na avaliação da maioria dos quesitos, o que demonstra que os leitores não perceberam diferenças significativas entre a notícia escrita pelo repórter e a notícia escrita pelo *software*. (Ibidem, p. 526)

Um segundo estudo (VAN DER KAA, KRAHMER, 2014) desenvolvido na Holanda analisou especificamente a percepção dos leitores quanto à credibilidade das notícias automatizadas de esportes e finanças avaliadas por 232 pessoas, sendo 64 delas jornalistas. Nesta pesquisa, os textos indicavam previamente se haviam sido escritos por um jornalista

ou por um computador, e os resultados apontam que os leitores avaliaram as notícias automatizadas mais confiáveis do que aquelas escritas por um repórter, enquanto os jornalistas se mostraram surpresos positivamente quanto a qualidade dos textos escritos por algoritmo, deixando observações como *“Isto na verdade não é ruim para um computador”*⁴⁶. (VAN DER KAA, KRAHMER, 2014, p. 3)

Figura 7 - Percepção dos leitores para notícias escritas por *software* e por jornalista



Fonte: CLERWALL, 2014, p. 525

As outras duas pesquisas foram conduzidas na Alemanha (GRAEFE et al., 2016; HAIM, GRAEFE, 2017) e ambas compararam notícias automatizadas com notícias escritas por jornalistas, sendo que a primeira contou com 986 participantes e teve o intuito de realizar um “Teste de *Turing*”⁴⁷ do jornalismo, e a segunda com 618 pessoas buscou avaliar a percepção dos leitores seguindo 13 quesitos em dois momentos diferentes: antes da leitura das notícias (expectativa) e após a leitura (percepção).

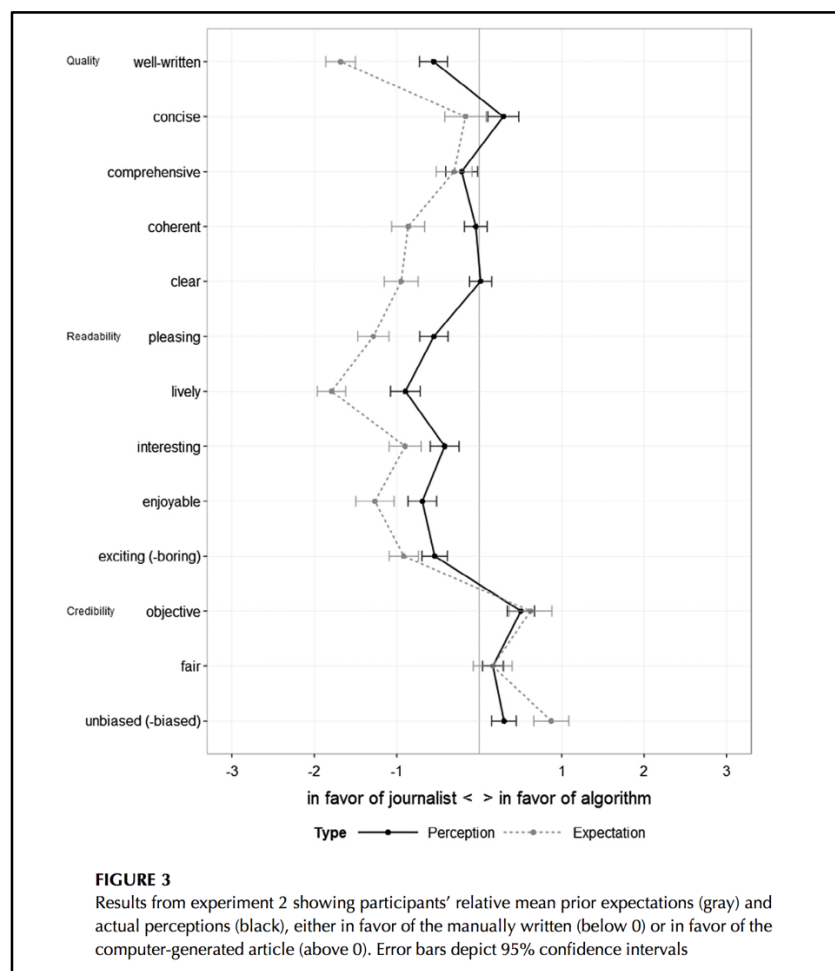
Em termos de avaliação do conteúdo, os dois estudos obtiveram resultados semelhantes que corroboram a pesquisa de Clerwall (2014), onde concluem que as notícias

⁴⁶ Texto original: “This is actually not bad for a computer”. (VAN DER KAA, KRAHMER, 2014, p. 3)

⁴⁷ O “Teste de *Turing*” foi um conceito introduzido por Alan Turing (1950) em seu artigo “Computing Machinery and Intelligence” e tinha como objetivo testar a capacidade de uma máquina em exibir comportamento inteligente equivalente a um ser humano, ou indistinguível deste.

automatizadas tendem a ser melhor classificadas em termos de credibilidade, enquanto as notícias escritas por jornalistas têm maior legibilidade. Os estudos também observaram que alguns leitores tiveram dificuldade de distinguir qual notícia havia sido escrita por um jornalista e qual fora produzida por um *software*. Para Graefe et al. (2016), uma provável explicação para esta boa avaliação é o fato das notícias automatizadas seguirem rigorosamente um *template* redigido por um jornalista, contendo assim todas as características convencionais da redação de uma notícia. (GRAEFE et al, 2016, p. 17)

Figura 8 - Leitores avaliam a expectativa e a percepção de notícias automatizadas em uma análise comparada com notícias escritas por um jornalista



Fonte: HAIM, GRAEFE, 2017, p. 10

Na segunda pesquisa conduzida na Alemanha, um fator que chama a atenção é a avaliação de que as notícias automatizadas são percebidas pelos leitores como menos preconceituosas, o que pode ser visualizado no gráfico da Figura 8, no último indicador *unbiased/biased*. Para Haim e Graefe (2017), este resultado sugere que o senso comum

acredita que os algoritmos são capazes de fornecer relatos mais imparciais sobre um fato/acontecimento.

A partir dos resultados destas pesquisas, é interessante destacar a preocupação dos pesquisadores quanto à recepção das notícias automatizadas por jornalistas e leitores, e um desejo de se avaliar se estes conteúdos têm sua qualidade reconhecida por estes públicos, o que legitimaria a sua produção. Apesar de apresentarem metodologias clássicas que não são muito comuns nos estudos de jornalismo, estas pesquisas reverberam uma inquietação que já estava presente desde os primeiros experimentos com previsões do tempo automatizadas. Uma das conclusões de Glahn (1970) é que seu sistema estava apto para produzir textos com qualidade suficiente para serem aceitos pelo público, mas uma das ressalvas feitas pelo pesquisador se atenta para que tecnologias como esta não sejam responsáveis por reduzir a autonomia dos profissionais.

3.4 Jornalismo Automatizado: um campo de estudos em formação

Apesar da automação já estar presente no jornalismo há algumas décadas, podemos afirmar que o Jornalismo Automatizado é um campo do conhecimento recente, alvo de pesquisas específicas sobre a implementação de *softwares* NLG para a redação de notícias somente a partir de 2012. Um dos exemplos que confirmam isso é o termo Jornalismo Automatizado, que só começou a ser adotado de forma massiva a partir de 2016, quando o *Guide to Automated Journalism* (GRAEFE, 2016) foi publicado pelo *Tow Center for Digital Journalism*, se tornando um marco desta “estabilização” conceitual. Antes disso, diferentes termos eram utilizados como “repórter robô” (*robot reporter*, CARLSON, 2014), “notícias escritas por máquina” (*machine-written news*, VAN DALEN, 2012), ou “Jornalismo Algorítmico” (*algorithmic journalism*, DÖRR, 2015), para citarmos alguns exemplos. Uma das evidências de que a denominação do campo como Jornalismo Automatizado tende a ser mantida para as próximas pesquisas é o uso recorrente desta terminologia nos artigos mais recentes, tendo inclusive sido adotada por Dörr (2017), que não faz mais uso de “Jornalismo Algorítmico”.

Neste cenário, apesar de cada pesquisa ter sua especificidade, todas elas buscam construir um novo campo de conhecimento com informações, análises e reflexões sobre a adoção desta nova tecnologia por empresas jornalísticas, e um dos desafios que se coloca é estabelecer uma definição sobre o que seria o Jornalismo Automatizado. Carlson (2014) o define como “processos algorítmicos que convertem dados em narrativas jornalísticas com

ação limitada ou nenhuma intervenção humana além das escolhas iniciais de programação”⁴⁸ (CARSLON, 2014, p. 1), enquanto Graefe (2016) conceitua como “processos com o uso de *softwares* e algoritmos que automaticamente geram notícias sem nenhuma intervenção humana – depois da etapa inicial de programação do algoritmo, claro”⁴⁹. (GRAEFE, 2016, p. 14) Apesar de semelhantes, estas definições possuem três pequenas diferenças que devem ser pontuadas. Enquanto Carlson (2014) cita apenas algoritmos, fala da conversão de dados em textos jornalísticos com ação limitada ou nenhuma intervenção humana, Graefe (2016) inclui em sua definição os *softwares* e enfatiza a não existência de nenhuma intervenção humana além da etapa inicial de programação, eliminando os processos de conversão de dados e a “ação limitada” dos humanos.

Propondo uma definição mais elaborada que considera a complexidade dos *softwares* NLG no campo da linguística computacional, Dörr (2015) reconhece que os algoritmos não são capazes de gerar textos sem a interferência humana, mas argumenta que

o elemento humano direto e ativo é eliminado do processo de criação no Jornalismo Algorítmico. Com isso, não quero dizer que o fator humano é eliminado da criação do conteúdo completamente, pois os algoritmos são desenvolvidos por humanos. Mas o ponto é que, com os *softwares* NLG, o papel do jornalista passa a ser mais indireto (NAPOLI, 2014) antes, durante e depois da produção do texto; por exemplo, seleção de fonte (*input*), verificação do fato, a escrita e a distribuição (*output*) são automatizadas e pré-selecionadas no Jornalismo Algorítmico. Agora, o trabalho e o conhecimento jornalísticos têm que incluir cada vez mais habilidades de programação.⁵⁰ (DÖRR, 2015, p. 9)

Nos incomoda o fato de todas estas três definições limitarem a ação humana na produção de uma notícia automatizada unicamente à etapa de programação dos algoritmos, uma visão um tanto quanto simplificada e limitada que não consegue se atentar para a complexidade que envolve o desenvolvimento, a implementação, o monitoramento e a manutenção em atividade de um *software* desta natureza. Um dos argumentos que defendemos nesta pesquisa é o de que, por trás dos processos de produção automatizada de uma notícia, há uma rede complexa envolvendo diversos atores humanos e não-humanos que não pode ser reduzida e nem simplificada a apenas uma rede de bancos de dados, *templates*,

⁴⁸ Texto original: “algorithmic processes that convert data into narratives news texts with limited to no human intervention beyond the initial programming choices.” (CARLSON, 2014, p. 1)

⁴⁹ Texto original: “Automated journalism refers to the process of using software or algorithms to automatically generate news stories without human intervention – after the initial programming of the algorithm, of course.” (GRAEFE, 2016, p. 14)

⁵⁰ Texto original: “I argue that the direct and active human element *during* the process of content creation is eliminated in Algorithmic Journalism. This is not to say that human factor is eliminated from content creation altogether, because algorithms are themselves developed by humans. The point is that the individual journalist in NLG is changing to a more indirect role (Napoli 2014) before, during and after text production; as, for example, source selection (input), fact checking, the actual writing (throughput – both depending on coding) and distribution (output) are automated and pre-selected in Algorithmic Journalism. Journalistic work and knowledge now has to comprise skills like programming even more.” (DÖRR, 2015, p. 9)

algoritmos e programadores. Esta visão, no nosso ponto de vista, reduz o campo das Ciências da Computação à tarefa única e exclusiva de programação de algoritmos, e não se atenta para as outras especificidades necessárias para o desenvolvimento de qualquer sistema.

Esta perspectiva conceitual tecnocêntrica pode ser melhor compreendida se aproximarmos a pesquisa sobre Jornalismo Automatizado de um campo maior, o do Jornalismo Computacional, que procura enfatizar os chamados “impactos” dos algoritmos e do *big data* nas práticas jornalísticas.

O termo Jornalismo Computacional⁵¹ surgiu em 2006 em um simpósio no *Georgia Institute of Technology* liderado pelo professor Irfan Essa que reuniu pesquisadores de Ciências da Computação e jornalistas em Atlanta, nos Estados Unidos, para discutir o desenvolvimento e a aplicação de *softwares* que auxiliassem nas tarefas de apuração, organização e divulgação de informações precisas e confiáveis na rotina das redações. Em 2009, um novo *workshop* no *Center for Advanced Study in the Behavioral Sciences* (CASBS), na *Stanford University*, também nos Estados Unidos, deu origem a um relatório que define o Jornalismo Computacional como uma área promissora e em plena expansão abrangendo a “combinação de algoritmos, dados, e conhecimento das ciências sociais para complementar o papel de regulação do jornalismo”⁵². (HAMILTON, TURNER, 2009, p. 2)

Este campo de estudos se estabelece a partir da associação de duas abordagens – a Reportagem Assistida por Computador (RAC) e o uso de ferramentas das ciências sociais pelo jornalismo – e propõe desenvolver sistemas que auxiliem os jornalistas a lidar com o crescente volume de dados estruturados e não-estruturados na produção de uma reportagem. (DIAKOPOULOS, 2010) A partir das experiências já desenvolvidas pela RAC, o Jornalismo Computacional seria o próximo passo para jornalistas e cientistas da computação desenvolverem novas ferramentas com a promessa de reduzir os custos e as dificuldades de se fazer uma reportagem aprofundada e investigativa sobre assuntos de interesse público. (COHEN et al., 2011, p. 66)

Se analisadas sob a perspectiva do STS, parece-nos que estas primeiras abordagens sobre o Jornalismo Computacional reverberam um ponto de vista determinista que credita à tecnologia o papel revolucionário de solucionar, através de ferramentas, todos os problemas acarretados pela crise econômica do jornalismo, creditando neste novo campo a salvação para as empresas continuarem investindo na produção de reportagens. Hamilton e Turner (2009) acreditam que estas novas ferramentas irão mudar a forma como lidamos com os

⁵¹ Disponível em: https://en.wikipedia.org/wiki/Computational_journalism Acessado em: 11/12/2017

⁵² Texto original: “combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism”. (HAMILTON, TURNER, 2009, p. 2)

dados, em um processo mais transparente onde as informações brutas utilizadas como fonte de uma reportagem seriam compartilhadas com os leitores. Adicionalmente, Diakopoulos (2011) prevê a necessidade das escolas de jornalismo ensinarem técnicas de modelagem, parametrização, agregação, escalabilidade e programação, vislumbrando oportunidades que parecem surgir para profissionais que desenvolverem novas formas de apuração, organização e divulgação de notícias. (Idem, p. 4)

Ao reconhecer as limitações destas primeiras pesquisas sobre Jornalismo Computacional, que se preocupam principalmente com o desenvolvimento de ferramentas digitais a serem adotadas pelos jornalistas, Anderson (2012) sugere uma nova abordagem sociológica para este campo na tentativa de abarcar as preocupações dos jornalistas quanto à implementação dos algoritmos nas redações, lançando um olhar para os impactos sociais, políticos, organizacionais e culturais ligados ao desenvolvimento destas novas tecnologias, que não estão imunes às redes de poder incorporadas ao jornalismo. Anderson (2012) chama a atenção para a natureza híbrida das tecnologias inseridas nas redações jornalísticas, e cita Latour (1993) para refletir que os algoritmos de mineração de dados e outros dispositivos cada vez mais invisíveis não são totalmente materiais, e nem inteiramente humanos, são entidades híbridas compostas de intencionalidade humana e obstáculos materiais, onde o humano se torna parcialmente inflexível e o material parcialmente intencional. (ANDERSON, 2012, p. 1016)

As discussões propostas, entre outros, por Gillespie (2014) e Lewis e Westlund (2014) em torno de dois conceitos-chave para o Jornalismo Automatizado – os algoritmos e o *big data* respectivamente – nos ajudam a situar uma abordagem que rompe com a visão tecnocêntrica do Jornalismo Computacional e apresenta uma argumentação mais centrada nas dinâmicas sociotécnicas e tecnopolíticas.

Como define Gillespie (2014), algoritmos são “procedimentos codificados que transformam o dado de um *input* em um *output* desejado” (GILLESPIE, 2014, p. 167). Eles são desenvolvidos para agirem de forma automática, sem a necessidade de intervenção humana, mas para isso seguem regras pré-determinadas e um banco de dados estruturado, o que significa que o *input* de dados teve que ser limpo e organizado para fornecer informações de forma precisa para os algoritmos poderem agir. O algoritmo não é um objeto com uma ação única, estável e delimitada, ele não age sozinho e sim faz parte de uma rede onde a combinação de vários algoritmos dá origem a um *software*, a uma página na internet,

a uma plataforma digital. “Na verdade, quando nos referimos a um algoritmo, normalmente não estamos falando de um, mas de muitos algoritmos.”⁵³ (GILLESPIE, 2014, p. 179)

Ao analisar como os algoritmos operam para tomar suas decisões, Diakopoulos (2014) os categoriza em quatro grupos – priorização, classificação, associação e filtragem – e discute como o poder algorítmico se estabelece a partir de regras articuladas por programadores e por técnicas de aprendizagem de máquina. (DIAKOPOULOS, 2014, p. 13) O pesquisador chama a atenção para a dimensão humana dos algoritmos que exercem influência nos critérios de escolha e interpretação dos dados.

A responsabilidade (*accountability*) algorítmica deve, portanto, considerar os algoritmos como objetos criados por humanos e levar em conta as intenções, de grupos ou processos institucionais, que influenciaram o design, assim como a agência de atores humanos na interpretação do *output* em processos superiores de tomada de decisões.⁵⁴ (DIAKOPOULOS, 2014, p. 5)

Diakopoulos (2014) é um dos pesquisadores que pontua a importância de se esclarecer quais tipos de dados estão sendo utilizados como *input*, quais os critérios estabelecidos para priorizar, classificar, associar e filtrar estas informações. “Considerando que as organizações jornalísticas irão empregar algoritmos nos processos de produção de notícias, seja para encontrar pautas a partir de gigantescos bancos de dados ou para criar narrativas interativas, padrões de transparência de algoritmos aplicados ao jornalismo precisarão ser desenvolvidos.”⁵⁵ (Idem, p. 13)

Gillespie (2014) argumenta que uma das dificuldades de se entender como os algoritmos operam reside no fato deles imbuírem uma falsa ideia de imparcialidade, sustentada pelo discurso de grandes empresas de tecnologia – como o *Facebook* e o *Google* –, que não revela as redes sociotécnicas por trás destas plataformas cada dia mais presentes nas nossas vidas. Estas empresas fomentam a ideia de neutralidade algorítmica como uma estratégia para não assumirem a responsabilidade pela ação dos algoritmos, cada vez mais complexos, que desenvolvem.

Em sua análise, Gillespie (2014) pontua que a objetividade algorítmica, em vários aspectos, opera de maneira análoga à objetividade jornalística. Semelhante às ferramentas de busca, jornalistas desenvolvem táticas para determinar quais assuntos são mais relevantes,

⁵³ Texto original: “In fact, what we might refer to as an algorithm is often no one algorithm but many.” (GILLESPIE, 2014, p. 179)

⁵⁴ Texto original: “Algorithmic accountability must therefore consider algorithms as objects of human creation and take into account intent, including that of any group or institutional processes that may have influenced their design, as well as the agency of human actors in interpreting the output of algorithms in the course of making higher-level decisions.” (DIAKOPOULOS, 2014, p. 5)

⁵⁵ Texto original: “As news organizations also come to employ algorithms in the shaping of the news they report, whether that be in finding news stories in massive datasets or presenting stories interactively, journalistic standards for transparency of algorithms will need to be developed.” (DIAKOPOULOS, 2014, p. 13)

como verificar a confiabilidade das informações, como elas devem ser reportadas para o leitor. Como as ações dos algoritmos, estas práticas são relativamente invisíveis e são seguidas criteriosamente pelos jornalistas na busca pela objetividade como princípio, mesmo que esta não possa ser alcançada tendo em vista a impossibilidade de se eliminar as subjetividades, os julgamentos de valor e de inclinação política. Para os jornalistas, a objetividade seria como um “ritual estratégico” (TUCHMAN, 1972 em GILLESPIE, 2014) que auxilia o profissional a realizar escolhas cuidadosas para alcançar um certo grau de legitimidade e se afastar dos próprios preconceitos e convicções. Por outro lado, a objetividade algorítmica está fundamentada na promessa de neutralidade mecânica da tecnologia, e não dependeria diretamente da expertise técnica ou de regras institucionais estabelecidas. Mas em ambas, a objetividade está ligada à busca por legitimidade a partir de diretrizes processuais que determinam como será realizada a seleção de informações. (GILLESPIE, 2014, p. 181)

Neste cenário, Gillespie (2014) afirma que os algoritmos introduzem uma nova lógica nas redações jornalísticas que pode concorrer com o trabalho executado pelos editores, onde a lógica editorial depende das escolhas subjetivas de profissionais treinados, autorizados institucionalmente e validados pelo público. De modo análogo, a lógica algorítmica depende de escolhas processuais desempenhadas por um computador, projetadas por programadores para automatizar algum *proxy* de julgamento humano ou descobrir padrões nos traços sociais coletados. A ação de ambos – editores e algoritmos – busca identificar informações relevantes para o público, livres de erros, tendências e manipulações, em abordagens importantes, mas igualmente problemáticas e cabíveis de armadilhas. (Idem, p. 192)

Um outro pilar de sustentação do Jornalismo Computacional é o *big data* e o esforço de se desenvolver sistemas que consigam processar enormes quantidades de dados gerando um *output* que agregue valor a estas informações.

Lewis e Westlund (2014) chamam a atenção para os vários aspectos sociais, culturais e tecnológicos em torno da ideia mística que a palavra *big data* acaba criando em seu entorno, em um esforço de encontrar caminhos para lidar com a abundância de dados, criar técnicas de análise e visualização, sem desconsiderar a mitologia que surge em torno dos *insights* guiados por dados, e as consequências destas práticas para a democracia e a sociedade. “Este momento ‘*big data*’ não é apenas uma transição tecnológica para o dilúvio de dados. Pelo contrário, este é um fenômeno sociotécnico com origens e implicações culturais, econômicas e políticas; é, na verdade, uma mitologia tanto quanto uma ciência e também um negócio.”⁵⁶ (LEWIS, WESTLUND, 2014, p. 2)

⁵⁶ Texto original: “This ‘big data moment’ is not merely a technological transition toward data deluge. Rather, it is a sociotechnical phenomenon with cultural, economic, and political origins and implications; it is, indeed, a mythology as much as a science or business.” (LEWIS, WESTLUND, 2014)

Os autores endossam que o *big data*, como qualquer outro dado, não representa uma verdade objetiva e absoluta, e desenvolvem quatro lentes conceituais estabelecidas nos campos da epistemologia, expertise, economia e ética para refletir sobre o potencial uso do *big data* pelas redações jornalísticas, reconhecendo neste momento o interesse jornalístico tanto do lado profissional quando do lado comercial de se buscar iniciativas que agreguem sentido e valor a este volume crescente de dados que tomou conta da sociedade.

Contrários à ideia de que o *big data* e a automatização possam substituir o trabalho humano, Lewis e Westlund (2014) acreditam na associação deste grande volume de dados à ideia de ampliação, denominada posteriormente por Linden (2017) e Marconi (2017) de Jornalismo Aumentado – *Augmented Journalism* em inglês -, o que catalisaria novas formas tecnológicas de se trabalhar com a notícia. Marconi (2017) descreve o jornalismo aumentado como o uso de “Inteligência Artificial para empoderar os jornalistas a criar reportagens em grandes escalas e com maior precisão”⁵⁷ (MARCONI, 2017, p. 2), e pontua três principais preocupações que devem ser levadas em conta neste contexto: os riscos inerentes à geração de notícias algorítmicas não verificadas, o potencial de interrupções do fluxo de trabalho e o crescente hiato nos conjuntos de habilidades necessárias para o gerenciamento desta nova especialidade. (Idem, p. 3)

Para Lewis e Westlund (2014), em um cenário de crescente automatização, os jornalistas se transformam em “gestores do conhecimento”, profissionais qualificados responsáveis por reunir, organizar, analisar os fluxos de informação em uma comunidade. Em uma rede heterogênea, jornalistas, analistas de dados e programadores serão os responsáveis por inscrever nos algoritmos a lógica e os valores jornalísticos.

⁵⁷ Texto original: “artificial intelligence to empower its journalists to report at greater scale and with greater accuracy.” (MARCONI, 2017 p. 2)

4 Jornalismo Automatizado: cartografando três estudos de caso

Neste capítulo, cartografamos três estudos de casos que exemplificam como veículos jornalísticos adotaram algoritmos e *softwares* NLG em suas redações. Buscamos inspiração no STS, na Teoria Ator-Rede e na Cartografia de Controvérsias com o intuito de criar descrições detalhadas de atores e de declarações vinculadas ao Jornalismo Automatizado que nos auxiliem a abrir as caixas-pretas e revelar como as notícias automatizadas são produzidas, traçando assim algumas das redes sociotécnicas que se vinculam a esta inovação tecnológica.

O primeiro estudo de caso descreve a experiência da *startup Narrative Science*, desde o projeto piloto desenvolvido na *Northwestern University*, passando pela polêmica afirmação em 2011 de um de seus fundadores de que, em cinco anos, um computador ganharia o *Pulitzer*, detalhando as experiências de produção automatizada de notícias com a *Forbes* e com a *ProPublica*, até o anúncio em 2016 de que eles não atuavam mais no mercado jornalístico.

O segundo estudo de caso apresenta a rede formada em torno de outra *startup* inicialmente denominada *StatSheet* e posteriormente *Automated Insights*, em um percurso semelhante ao da *Narrative Science*. Um início marcado pelo lançamento de 345 sites para a cobertura automatizada de jogos de basquete da liga universitária dos Estados Unidos que avança para uma parceria com a agência *Associated Press* com o objetivo de produzir notícias automatizadas sobre os resultados financeiros de empresas listadas na bolsa de valores.

O terceiro estudo de caso tem como foco a adoção do Jornalismo Automatizado pelo *Los Angeles Times* e descreve dois projetos desenvolvidos pela equipe interna de jornalistas de dados: um para a cobertura de homicídios – o *The Homicide Report* – e outro para terremotos – o *Quakebot*.

Ao final, evidenciamos algumas semelhanças e especificidades destes estudos de caso em busca de um olhar mais reflexivo que nos ajude a compreender o contexto no qual o Jornalismo Automatizado é introduzido na rotina dos jornalistas e das redações, e como esta inovação dialoga com o crescente processo de digitalização das práticas jornalísticas.

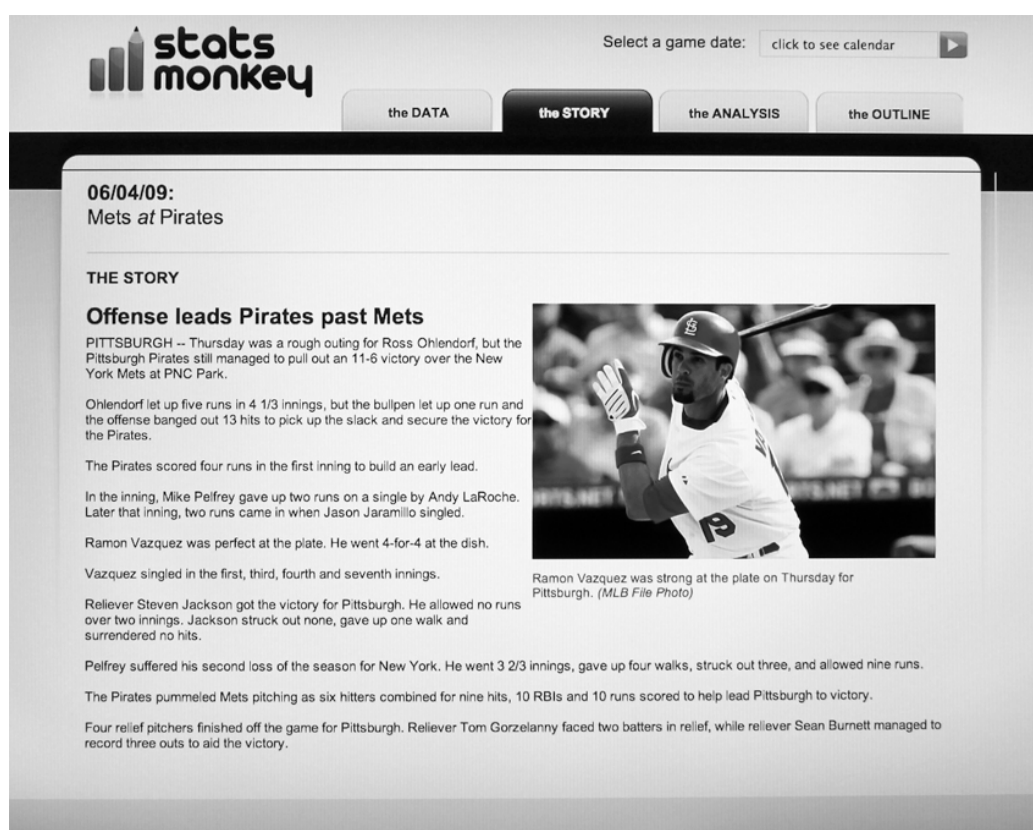
4.1 *Narrative Science*: do *Pulitzer* à desistência do jornalismo

4.1.1 *StatsMonkey*

No primeiro semestre de 2009, uma disciplina de pós-graduação reuniu jornalistas e programadores na *Northwestern University*, nos Estados Unidos, com o objetivo de criar

sistemas que transformassem dados em histórias. A proposta desafiadora – intitulada *Interactive Innovation Project* – era liderada por dois professores de jornalismo – Rich Gordon e Jeremy Gilbert – e dois professores de ciências da computação - Kristian Hammond e Larry Birnbaum. (GORDON, 2009) No final da disciplina, foi realizada uma apresentação dos cinco projetos desenvolvidos para uma plateia que incluía, além dos estudantes e professores, jornalistas locais e executivos da *ESPN*, do *The Tribune* e de empresas de tecnologia. E um dos projetos chamou particularmente a atenção de todos: o *StatsMonkey*. (LEVY, 2012)

Figura 9 - Exemplo de notícia automatizada escrita pelo *StatsMonkey*



Fonte: BROWN, 2015

Ainda um protótipo com várias limitações, a proposta do *StatsMonkey* era escrever de forma automatizada notícias de jogos de beisebol da liga universitária utilizando estatísticas disponíveis na internet. O banco de dados utilizado pelo *software* era composto basicamente dos resultados finais dos jogos, estatísticas dos principais lances e informações sobre o campeonato. (ALLEN et al, 2010)

O projeto foi desenvolvido por uma equipe multidisciplinar que incluía o jornalista John Templon, que tinha experiência prévia na cobertura de jogos esportivos universitários, o jornalista-programador Nicholas Allen, o jornalista Tian Huang e a aluna de computação e ciências da informação Thu Cung. (GORDON, 2009) Em uma notícia publicada na época,

John Templon relata quais foram os maiores aprendizados daquele curso de pós-graduação: *“O que eu vou levar comigo destas aulas é o que é preciso para trabalhar em uma equipe (multidisciplinar), alguma compreensão sobre a filosofia por trás do Agile design e um pouco de conhecimento técnico sobre Python.”*⁵⁸ (LEOPOLD, 2009)

Apesar de ser um projeto interno desenvolvido em uma universidade, o *StatsMonkey* acabou atraindo a atenção dos jornalistas na época por conta de uma notícia publicada na *Medill Magazine*, revista vinculada à escola de jornalismo da *Northwestern University*. Na reportagem, os próprios estudantes explicam o que os motivou a desenvolver o *StatsMonkey*. Para Nicholas Allen, esta tecnologia tem potencial para ampliar a cobertura de notícias esportivas: *“Há uma enorme quantidade de eventos esportivos que não são cobertos, principalmente em nível universitário. Então, pensamos que nossa tecnologia tem um grande potencial neste espaço.”*⁵⁹ (GREINER, 2009) Já Tian Huang revela um desejo de que esta tecnologia libere jornalistas de tarefas repetitivas: *“Nosso projeto é uma maneira de manter o jornalismo próspero. Ao expandir a cobertura e liberar os repórteres de escrever as histórias básicas de um jogo, estamos dando aos jornalistas tempo de se concentrarem nos detalhes e nas análises que levam os leitores a suas publicações.”*⁶⁰ (Idem)

A divulgação do *StatsMonkey* acabou gerando reações, nem sempre positivas, para esta tecnologia, acendendo um alerta entre os jornalistas, como Andrew Greiner, da *NBC Chicago*, que afirmou que *“Em breve, os repórteres de esporte devem se tornar obsoletos”*⁶¹ (GORDON, 2009), ou Rick Green do *The Hartford Courant*: *“[...] não se perde algo quando o repórter não está nos jogos, observando os jogadores, prestando atenção no que não é dito e sentindo o clima?”*⁶² (Ibidem) Outro jornalista que também se pronunciou na época foi David Carr para o jornal *The New York Times*: *“O estranho sobre o StatsMonkey é o fato das notícias não serem terríveis”*⁶³ (CARR, 2009); o que também chamou a atenção da jornalista Cathal Kelly do *Toronto Star*: *“Não há nenhum erro aparente de apuração ou de gramática. Escritores reais cometem dezenas, ou talvez centenas, de erros. StatsMonkey, um objeto sintético e sem*

⁵⁸ Texto original: “What I'll take away from the class is what it takes to work in a (multidisciplinary) team, some understanding about Agile design philosophy and a little more technical knowledge of things Python.” (LEOPOLD, 2009)

⁵⁹ Texto original: “There are a tremendous number of sporting events that go uncovered, even at the college level. So we think that there is great potential for our product in that space.” (GREINER, 2009)

⁶⁰ Texto original: “Our project is a way to keep journalism thriving. By expanding coverage and freeing reporters from writing the basic game stories, we're giving journalists the time to really focus on the features and analysis that drive readers to their publications.” (Idem)

⁶¹ Texto original: “Soon enough, sports reporters could be obsolete” (GORDON, 2009)

⁶² Texto original: “... isn't something lost when the reporter isn't there at the games, talking to players, paying attention to what's not said and feeling the mood?” (Idem)

⁶³ Texto original: “The weird thing about StatsMonkey is how not-that-terrible the stories are”. (CARR, 2009)

coração, tem a intenção de colocar todos os revisores na fila para receberem ajuda do governo.”⁶⁴ (KELLY, 2009)

Em um tom mais otimista, a editora associada da *ESPNChicago.com* Kristin Lanning aponta que esta tecnologia pode ser a salvação para que a cobertura de jogos universitários não deixe de ser realizada.

Em um momento onde os repórteres que cobrem os jogos universitários estão (tragicamente) sendo demitidos, a página sobre a cobertura destes jogos corre o risco de ser extinta. Esta nova tecnologia pode dar aos esportes universitários a cobertura online que eles merecem. Você ficaria impressionado se soubesse quantas pessoas clicam nesta página. Os esportes universitários devem ser valorizados pelo jornalismo esportivo, e eu agradeço qualquer tecnologia que mantenha esta página no ar.⁶⁵ (MEDILL MAGAZINE, Winter 2009, p. 5)

Por outro lado, o jornalista Doug Bartholomew abusa da ironia para sugerir que *softwares* como o *StatsMonkey* podem um dia substituir o trabalho dos editores.

Com as notícias esportivas geradas por máquina da Medill's – oops, desculpe, o *StatsMonkey* – em mente, eu prevejo o dia não muito distante no futuro quando os editores vão comemorar o lançamento do *Managing Editor 2.1* – oh, desculpe, o nome real é Suporte de Decisão Editorial (SDE). Enquanto a versão anterior SDE 2.0 não atingia uma profundidade suficiente, a última versão do *software* promete finalmente eliminar as tarefas ‘repetitivas’ e ‘chatas’ que controlam os editores, a saber, avaliar as notícias todos os dias, se reunir com os subeditores, editar e revisar as principais notícias.⁶⁶ (Ibidem, p. 5)

Para acalmar os ânimos, o professor da *Northwestern University* Rich Gordon, um dos responsáveis pela disciplina na qual o *StatsMonkey* foi desenvolvido, também se pronunciou sobre o assunto. “*Eu não acho que os jornalistas esportivos devam se preocupar – se o StatsMonkey se tornar um produto comercial, é altamente improvável que os jornalistas fiquem sem trabalho.*”⁶⁷ (GORDON, 2009) Gordon (2009) destaca a ação limitada do programa e a baixa qualidade dos textos, que não substituiriam uma notícia rica e bem estruturada escrita por um repórter profissional. Em seu artigo, ele lista algumas

⁶⁴ Texto original: “No discernible errors of fact or grammar. Real writers make dozens, possibly hundreds, of mistakes. StatsMonkey, the heartless synthetic is intent on putting all copy editors on the bread lines.” (KELLY, 2009)

⁶⁵ Texto original: “In a time where prep reporters are (tragically) being laid off, and the preps page itself is in danger of extinction, this new technology might be able to give high school sports the Internet coverage they deserve. You’d be amazed how many people click on the preps page. High school sports should always matter in sports journalism, and I welcome any technology that keeps the preps page going online.” (MEDILL MAGAZINE, Winter 2009, p.5)

⁶⁶ Texto original: “With Medill’s ‘Machine-Generated Sports Stories’ — oops, sorry, StatsMonkey — in mind, I foresee a day not too far in the future when publishers will cheer the release of Managing Editor 2.1 — oh, sorry, real name Editorial Decision Support (EDS). While its predecessor EDS 2.0 lacked sufficient in-depth functionality, the latest version of the software package promises to finally eliminate those ‘repetitive’ and ‘boring’ tasks managing editors face, namely, evaluating the news each day, meeting with department heads, and assigning and editing top stories.” (MEDILL MAGAZINE, Winter 2009, p.5)

⁶⁷ Texto original: “I don’t think sportswriters need to be worried — if StatsMonkey becomes a commercial product, it is highly unlikely to put sports journalists out of work.” (GORDON, 2009)

possibilidades do *StatsMonkey* auxiliar nas tarefas desempenhadas pelos jornalistas, como publicar instantaneamente uma notícia logo após o término do jogo, liberando o repórter para ir ao campo ou ao vestiário entrevistar os jogadores; produzir notícias sobre jogos de ligas menores, como campeonatos universitários, que normalmente não são pautados por jornalistas profissionais; ou ainda criar notícias específicas com foco nas estatísticas e no desempenho de um determinado jogador. (Idem)

4.1.2 Narrative Science

Uma das pessoas que se surpreendeu com a apresentação do *StatsMonkey* no final da disciplina da *Northwestern University* foi Stuart Frankel, um executivo da área de tecnologia que, na época, havia acabado de vender a sua empresa - a *DoubleClick* – para o *Google*. “Quando estes caras fizeram a apresentação, o clima da sala mudou. Mas ainda era apenas um pequeno software que escrevia notícias sobre jogos de beisebol – muito limitado.”⁶⁸ (LEVY, 2012)

Na busca por novos projetos, Frankel propôs investir para dar continuidade ao *StatsMonkey*, ideia que foi aceita por dois dos quatro estudantes – John Templon e Nicholas Allen – e os dois professores de ciências da computação – Kristian Hammond e Larry Birnbaum. Em um primeiro momento, os jornalistas John Templon e Nicholas Allen foram incorporados à *Northwestern University* como pesquisadores associados para aperfeiçoarem o *software* de automatização de notícias, período que durou entre junho de 2009 e abril de 2010, quando oficialmente a *Narrative Science* foi fundada.

Logo em seguida, a *startup* começou a trabalhar para seu primeiro cliente, a *Big Ten Network* [<http://www.btn.com/>] – uma *joint venture* entre a *Big Ten Conference* e a *Fox Networks* – onde implementou o *software* para produzir breves notícias sobre jogos de *baseball* e *softball* a partir das estatísticas e resultados. (LOHR, 2011)

Porém, um problema logo apareceu: as notícias davam maior destaque às vitórias, e acabavam exaltando times de fora da *Big Ten Conference* quando estes ganhavam as partidas, criando textos com um tom muitas vezes humilhante. Frente a este desafio, os engenheiros da *Narrative Science* alteraram alguns parâmetros do *software* para ajustar os ângulos da notícia e, em caso de derrota, enaltecer as jogadas heroicas e ignorar os erros dos times da *Big Ten*. (LEVY, 2012)

⁶⁸ Texto original: “When these guys did the presentation, the air in the room changed. But it was still just a piece of software that wrote stories about baseball games – very limited.”

Alguns meses depois, o *software* também foi adotado para a cobertura de jogos de basquete e futebol americano e, entre alguns resultados, observou-se uma melhora na classificação do conteúdo pelo *ranking* do *Google* e 40% de crescimento nos acessos se comparado com a audiência do site em 2009. (LOHR, 2011)

Em setembro de 2011, em uma reportagem publicada pelo *The New York Times*, Stuart Frankel e Kristian Hammond argumentam que a tecnologia que estavam desenvolvendo na *Narrative Science* poderia se tornar uma ferramenta barata para as empresas jornalísticas aumentarem o alcance de sua cobertura. (LOHR, 2011) Mas a grande repercussão desta notícia se deve a uma declaração em que Hammond enfatiza que “*em cinco anos, um programa de computador vai ganhar o Pulitzer, e eu serei condenado se não for a nossa tecnologia*”⁶⁹. (Idem)

Esta afirmação polêmica repercutiu em respostas de vários jornalistas, já que o *Pulitzer* é o prêmio de jornalismo mais prestigiado do mundo, um símbolo de qualidade e credibilidade, e esta afirmação, mais do que apenas chamar a atenção, parecia querer conferir legitimidade ao Jornalismo Automatizado. Uma destas reações foi de Evgeny Morozov com o artigo *A robot stole my Pulitzer!*, publicado em março de 2012 e traduzido para diversas línguas, inclusive português. Nele, o autor destaca a ironia por trás da tecnologia envolvendo o Jornalismo Automatizado, mais especificamente sobre as notícias sobre balancetes financeiros:

Não perca a ironia aqui: Plataformas automatizadas estão agora ‘escrevendo’ notícias sobre empresas que ganham dinheiro em negociações automatizadas. Esses relatórios são eventualmente encaminhados ao sistema financeiro, auxiliando outros algoritmos a detectar negócios ainda mais lucrativos. Essencialmente, isto é jornalismo feito por robôs e para robôs. O único ponto positivo aqui é que todo o dinheiro fica com humanos.⁷⁰ (MOROZOV, 2012)

Morozov (2012) ressalta como míope a ideia de que uma maior automação possa salvar o jornalismo, e defende o uso restrito desta tecnologia com a garantia da manutenção do emprego dos jornalistas, desde que esta inovação seja usada para que estes possam buscar projetos analíticos mais interessantes ao invés de reescreverem a mesma história todas as semanas. Para ele, a “*verdadeira ameaça vem da negação de não se investigar as consequências sociais e políticas de se viver em um mundo onde é quase impossível ler de*

⁶⁹ Texto original: “In five years, a computer program will win a Pulitzer Prize — and I’ll be damned if it’s not our technology.” (LOHR, 2011)

⁷⁰ Texto original: “Don’t miss the irony here: Automated platforms are now “writing” news reports about companies that make their money from automated trading. These reports are eventually fed back into the financial system, helping the algorithms to spot even more lucrative deals. Essentially, this is journalism done by robots and for robots. The only upside here is that humans get to keep all the cash.” (MOROZOV, 2012)

*forma anônima*⁷¹ (Idem) Ele destaca a dificuldade de se cultivar um pensamento crítico e não convencional em um mundo dominado por anunciantes e por gigantes da tecnologia como o Google, o Facebook e a Amazon. (Idem)

Uma outra jornalista a repercutir a polêmica em torno do Prêmio *Pulitzer* é Rebecca Greenfield que, para uma reportagem publicada pela *The Atlantic*, afirma que o mais assustador é a visão que Kristian Hammond tem sobre o que é jornalismo.

Em última análise, se os jornalistas (e as pessoas que dependem deles) devem ou não ter medo de um futuro maquínico vai depender do que você pensa que o jornalismo é. Há empresas inteiras construídas com base na ideia de produzir grandes quantidades de notícias, com a qualidade controlada por fórmulas semelhantes às de uma máquina. *Narrative Science* pode um dia colocar muitos desses jornalistas fora do trabalho. Mas quando a maioria das pessoas fala sobre jornalismo, eles não estão pensando em relatórios financeiros ou notícias sobre jogos de baseball. (Certamente, ninguém entra no jornalismo por uma paixão por tais coisas.) E encolher uma parte da indústria nunca será bom para a força de trabalho dela como um todo.⁷² (GREENFIELD, 2012)

Na mesma época da afirmação ao *The New York Times* sobre o Prêmio *Pulitzer*, a *Narrative Science* estava envolvida em um projeto com a *startup Game Changer* – um aplicativo que cria notícias de jogos estudantis de beisebol e basquete a partir de dados estatísticos fornecidos de forma colaborativa pelo público, em sua maioria técnicos, pais, amigos e familiares dos jogadores. (EMPSON, 2011) Entre algumas estatísticas, em 2011 o *Game Changer* produziu cerca de 400 mil textos, e a expectativa era de que este número chegaria a 1,5 milhões em 2012. (LEVY, 2012) Para Larry Birnbaum, gerar notícias automatizadas sobre jogos de pequenas ligas como as do *GameChanger* seria um esforço para facilitar e personalizar cada vez mais o acesso a informação.

Nós percebemos que ninguém quer 100 resultados de uma pesquisa, e nem mesmo 10. Você quer um resultado, algo que sintetiza as informações que você precisa agora. Queremos tornar a Inteligência Artificial contextualmente mais consciente. Em vez de sua pesquisa, o sistema empurra a informação diretamente para você, de uma forma que você pode usar.⁷³ (MCCORMICK MAG, Fall 2011, p. 24)

⁷¹ Texto original: "The real threat comes from our refusal to investigate the social and political consequences of living in a world where reading anonymously becomes a near impossibility" (MOROZOV, 2012)

⁷² Texto original: "Ultimately, whether or not journalists (and the people who rely on them) should fear a machine-powered future of journalism depends on what you think journalism is. There are whole businesses built on the idea of producing massive quantities of news stories, quality controlled by machine-like formulas. Narrative Science may one day put a lot of these journalists out of work. But when most people talk about journalism, they're not thinking about rote earnings reports or baseball game recaps. (Certainly no one goes into journalism out of a passion for such things.) And shrinking one part of an industry is never good for the workforce in the rest of it." (GREENFIELD, 2012)

⁷³ Texto original: "We realized you don't want 100 search results or even 10, You want one result, something that synthesizes the information you need right now. We want to make artificial intelligence more contextually aware. Instead of your searching, the system pushes the information directly to you, in a form you can use." (MCCORMICK MAG, Fall 2011, p. 24)

Um outro cliente da *Narrative Science* na época era um site de imóveis da *Hanley Wood*, e o presidente da empresa Andrew Reid, um ex-executivo da *Thomson Reuters*, exalta na mesma reportagem do *The New York Times* o custo-benefício da produção de conteúdos utilizando o *software*, pelo qual paga menos de US\$ 10 para um texto com aproximadamente 500 palavras. (LOHR, 2011)

4.1.3 Uma parceria com a *Forbes*

Alguns meses depois, em março de 2012, a *Narrative Science* anunciou uma parceria com a *Forbes* para a publicação automatizada de notícias sobre o mercado financeiro. As notícias produzidas pelo *Quill*, nome dado ao *software* desenvolvido pela *Narrative Science*, foram publicadas em uma seção criada exclusivamente para os balancetes financeiros automatizados, que pode ser acessada no link <http://www.forbes.com/sites/narrativescience>.

No lugar do nome do jornalista, estas notícias são assinadas pela *Narrative Science* e, ao final do texto, aparece informações dizendo que os dados foram fornecidos pela *Zacks Investment Research* e que a *Narrative Science* é uma plataforma que transforma dados em notícias e tendências, como é possível observar na Figura 10 de uma notícia automatizada publicada em julho de 2012.

Figura 10 - Notícia automatizada publicada na *Forbes*

Fonte: Reprodução de internet⁷⁴

⁷⁴ Disponível em: <https://www.forbes.com/sites/narrativescience/2012/07/19/forbes-earnings-preview-apple-3/#78c6e8106d9e>. Acessado em 08/09/2017

Semelhante às notícias escritas por um jornalista, o link “*Full Bio*” logo abaixo da assinatura da *Narrative Science* abre uma janela com uma descrição, *links* das notícias mais recentes e das mais acessadas assinadas por este autor. Nesta descrição, a *Forbes* retrata a *Narrative Science* como

líder na geração automatizada de narrativas para empresas. Desenvolvido com Inteligência Artificial, sua plataforma *Quill*™ analisa dados de diferentes fontes, entende o que é importante para o usuário final e, em seguida, gera automaticamente narrativas perfeitamente escritas para transmitir o significado dos dados para qualquer público consumidor ou comercial, em uma escala ilimitada. (forbes.com/sites/narrativescience)⁷⁵

Em uma entrevista para a revista *The Atlantic*, Lewis D'Vorkin, diretor de produto da *Forbes*, explica que o *Quill* é utilizado para escrever o que ele chama de “*previsões de lucro de empresas geradas por computador*”⁷⁶. A plataforma escolhe, ao processar os dados mais recentes de ações negociadas no mercado financeiro, quais empresas tiveram uma performance de destaque que mereça a publicação de uma notícia. (FASSLER, 2012) “*Com a Narrative Science, cobrimos mais de 600 notícias a cada temporada de resultados, enquanto antes tínhamos que escolher quem iríamos cobrir*”⁷⁷ (NARRATIVE SCIENCE, 2013), afirmou D'Vorkin em uma entrevista para o próprio site da *startup*. “*De fato, tivemos alguns casos onde previsões geradas pela Narrative Science foram as notícias mais lidas em nosso site.*”⁷⁸ (Idem) Em outra entrevista para a revista *Wired*, D'Vorkin revela que ficou impressionado, mas não surpreso, de não ter recebido nenhuma reclamação referente ao conteúdo das notícias automatizadas. (LEVY, 2012)

Na mesma reportagem para a revista *The Atlantic*, Stuart Frankel assume um discurso mais mercadológico e promete que a tecnologia da *Narrative Science* pode escrever sobre qualquer assunto usando qualquer tipo de dado, sugerindo ainda o surgimento de um novo profissional - o “*meta-jornalista*” – que ele descreve como “*peçoas que podem escrever milhões de notícias ao invés de uma única notícia de cada vez*”⁷⁹. (FASSLER, 2012) Para Frankel, seria só uma questão de tempo para o surgimento de mais notícias baseadas em dados estatísticos escritas por computador, e provoca: “*Mas existem tantas histórias para*

⁷⁵ Texto original: “leader in automated narrative generation for the enterprise. Powered by artificial intelligence, its *Quill*™ platform analyzes data from disparate sources, understands what is important to the end user and then automatically generates perfectly written narratives to convey meaning from the data for any intended consumer or business audience, at unlimited scale.” (forbes.com/sites/narrativescience)

⁷⁶ Texto original: “computer-generated company earnings previews”.

⁷⁷ Texto original: “With Narrative Science, we cover more than 600 stories each earnings season whereas before we had to pick and choose who we wanted to cover.” (NARRATIVE SCIENCE, 2013)

⁷⁸ Texto original: “In fact, there have been instances when previews generated by Narrative Science are the most read articles on our site.” (Idem)

⁷⁹ Texto original: “people who can write millions of stories opposed to a single story at a time”. (FASSLER, 2012)

*serem contadas que não são dirigidas por dados. São nelas que os jornalistas deveriam focar, não?*⁸⁰ (Ibidem)

Tanto Frankel quanto Hammond, também entrevistado pela *The Atlantic*, concordam que a tecnologia desenvolvida pela *Narrative Science* irá substituir alguns tipos de textos escritos por humanos, porém eles afirmam que o que mais os motiva é criar notícias raramente cobertas por jornalistas. Eles defendem o conceito de “*audiência para um*” onde sua plataforma criaria textos para serem lidos por uma ou poucas pessoas, no lugar de tabelas e dados confusos. (Ibidem) Hammond argumenta que os dados são extremamente valiosos, mas não no formato de planilhas e tabelas, e alega que onde há números, deveriam haver narrativas. Para ele, as pessoas não têm facilidade para interpretar números e um *software* NLG permite transformar dados numéricos e gráficos em narrativas facilmente compreendidas por qualquer um. “*A longo prazo, nossa tecnologia acaba por ser um mediador entre os dados e a experiência humana.*”⁸¹ (Ibidem)

Em uma outra reportagem que repercute a parceria da *Narrative Science* com a *Forbes*, Hammond dá mais uma declaração polêmica quando prevê que, em quinze anos, mais de 90% das notícias seriam escritas por um computador, mas em tom otimista ele afirma que o “*tsunami de robônotícias*” não irá eliminar o emprego de ninguém, imaginando um futuro onde, ao invés disso, o universo das notícias irá se “*expandir dramaticamente*”. (LEVY, 2012) Neste cenário, ele acredita que os *softwares* NLG se tornarão uma ferramenta para os jornalistas:

Talvez em algum momento, humanos e algoritmos colaborem, cada um agindo com a sua força. Os computadores, com suas memórias sem falhas e com a capacidade de acessar dados, podem atuar como legistas para escritores humanos. Por outro lado, os repórteres humanos podem realizar entrevistas e se focar nos detalhes – para depois enviá-los para um computador que escreverá tudo.⁸² (Idem)

Na mesma reportagem, Hammond conta o caso de uma festa onde encontrou uma pessoa incomodada com a tecnologia que estava desenvolvendo e o perguntou se ela seria responsável pela substituição de jornalistas por robôs. A sua resposta foi: “*Você já viu um repórter em um jogo da liga universitária? Esta é a coisa mais importante sobre nós. Ninguém perdeu nenhum emprego sequer por causa de nós.*”⁸³ (Ibidem)

⁸⁰ Texto original: “But there are so many stories to be told that are not data-driven. That’s what journalists should focus on, right?” (FRANKEL in FASSLER, 2012)

⁸¹ Texto original: “In the long run, our technology ends up being the mediator between data and the human experience.” (HAMMOND in FASSLER, 2012)

⁸² Texto original: “Maybe at some point, humans and algorithms will collaborate, with each partner playing to its strength. Computers, with their flawless memories and ability to access data, might act as legmen to human writers. Or vice versa, human reporters might interview subjects and pick up stray details—and then send them to a computer that writes it all up.” (LEVY, 2012)

⁸³ Texto original: “Have you ever seen a reporter at a Little League game? That’s the most important thing about us. Nobody has lost a single job because of us.” (LEVY, 2012)

Um dos jornalistas que se dedicou a pesquisar sobre o assunto, Joe Fassler revela ao final da sua reportagem para a *The Atlantic* que chegou à sede da *Narrative Science* com a expectativa de ter suas próprias convicções confirmadas: a de que a “*mente humana é um mistério sagrado*” e que a relação que os escritores e jornalistas têm com as palavras é única e profunda, duvidando da capacidade de escrever de um autômato.

Mas conversando com Hammond, eu percebi o quanto o processo da escrita – o que eu acreditava que era imprevisível, até variável – pode ser quantificado e modelado. Quando eu escrevo uma pequena história, eu estou fazendo exatamente o que estas plataformas de escrita fazem – usando a riqueza de dados (a minha experiência de vida) para fazer inferências sobre o mundo, apresentando estas inferências com um ângulo (ou tema), e criando uma estrutura adequada (baseada nos possíveis resultados que eu internalizei ao ler e observar, e ao fazer aulas de escrita criativa). É possível também dar a uma máquina uma cadência literária: escolher verbos fortes, substantivos específicos, deixar os advérbios longe, e assim por diante. Tenho certeza que algum gramático experiente poderia traçar todas as formas diferentes de tornar uma frase agradável (certamente, os oradores clássicos fizeram, com os seus quiasmos e *epanalepsis*, anáforas e antístrofes).⁸⁴ (FASSLER, 2012)

4.1.4 Projeto educacional com a *ProPublica*

Em 2011, a *ProPublica* lançou o projeto *The Opportunity Gap* [<http://projects.propublica.org/schools>] que reunia informações e gráficos sobre as escolas públicas e o desempenho dos estudantes nos Estados Unidos entre 2009 e 2010 a partir de dados fornecidos pelo *U.S. Department of Education Office of Civil Rights*. A premissa deste projeto era o questionamento “*O seu estado oferece acesso igualitário à educação?*”⁸⁵ e pretendia comparar o acesso a cursos avançados e programas de ensino em escolas localizadas em regiões mais ricas e mais pobres.

Com tabelas e informações que poderiam ser visualizadas em nível federal, estadual, distrital e específico de cada escola, o site permite comparar o desempenho escolar e o acesso a programas avançados de ensino associado a outros indicadores como raça, gênero e renda dos estudantes. Uma das conclusões do projeto foi que, em alguns estados, escolas

⁸⁴ Texto original: “But speaking with Hammond, I realized how much of the writing process—what I tend to think of as unpredictable, even baffling—can be quantified and modeled. When I write a short story, I’m doing exactly what the authoring platform does—using a wealth of data (my life experiences) to make inferences about the world, providing those inferences with an angle (or theme), the creating a suitable structure (based on possible outcomes I’ve internalized from reading and observing and taking creative writing classes). It’s possible to give a machine a literary cadence, too: choose strong verbs, specific nouns, stay away from adverbs, and so on. I’m sure some expert grammarian could map out all the many different ways to make a sentence pleasing (certainly, the classical orators did, with their chiasmus and epanalepsis, anaphora and antistrophe).” (FASSLER, 2012)

⁸⁵ Texto original: “Is Your State Providing Equal Access to Education?” Disponível em: <https://projects.propublica.org/schools/> Acessado em: 08/10/2017

em regiões mais pobres e com alto índice de estudantes de baixa-renda não tinham o mesmo acesso aos programas avançados de ensino. (LAFLEUR, 2011)

Dois anos depois, em 2013, a *ProPublica* resolveu atualizar as informações a partir de uma parceria com a *startup Narrative Science*, com o intuito de escrever uma notícia para cada uma das 52 mil escolas incluídas no projeto, a partir da análise dos dados e comparação dos resultados com a média na região onde aquela escola está localizada.

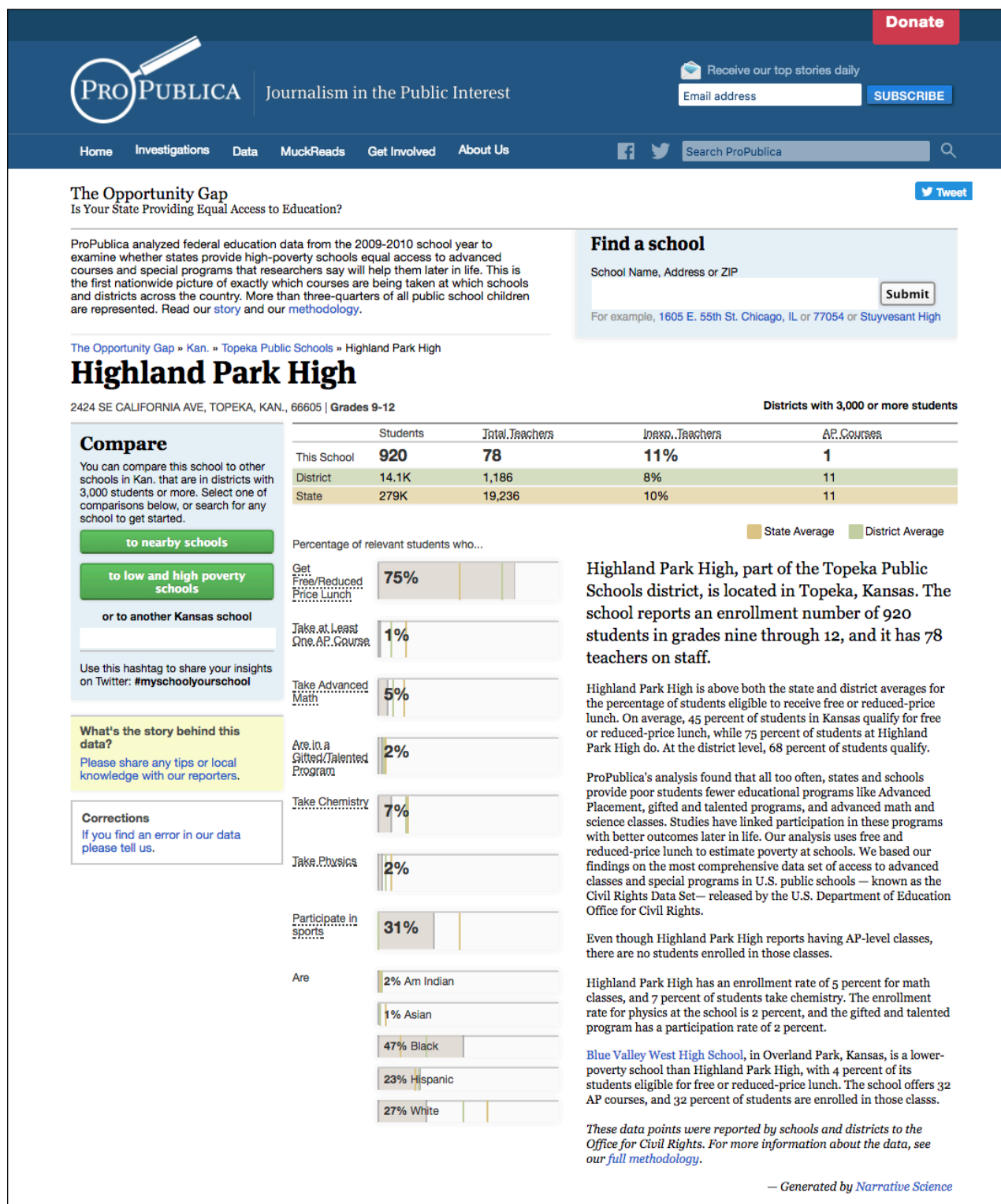
Na Figura 11, é possível ver uma destas notícias escritas pelo *software Quill*, da *Narrative Science*, sobre uma escola de Ensino Médio localizada no distrito de Topeka, no Kansas, onde há um alto índice de estudantes negros e latinos, sendo que 75% dos alunos recebem subsídio para o almoço. Os indicadores desta escola são abaixo da média tanto do distrito quanto do estado em termos de cursos avançados de matemática, física, química e esportes. A notícia automatizada oferece inicialmente um resumo dos dados específicos da escola e, posteriormente, compara estes resultados com o cenário no distrito onde ela está localizada e com a média alcançada pelo estado.

Um dos editores envolvidos neste projeto, Scott Klein comenta que a *ProPublica* foi apresentada à *Narrative Science* por um de seus investidores - *MacArthur Foundation* – e o projeto de atualização do *The Opportunity Gap* surgiu do interesse de ambas as empresas de transformar dados em histórias. (KLEIN, 2013) Para produzir os textos, os editores da *ProPublica* forneceram à equipe da *Narrative Science* acesso a todo o banco de dados do projeto e também alguns exemplos de notícias sobre algumas das 52 mil escolas. (LECOMPTE, 2015)

Para Klein, um dos maiores desafios foi editar ao mesmo tempo 52 mil narrativas, um processo que demonstrou para ele que o sistema desenvolvido pela *Narrative Science* é muito mais complexo do que apenas um texto com lacunas em branco e algumas variáveis interpoladas. “*Editar uma notícia não significa que você editou todas elas. (...) as edições que faziam sentido em um caso não fizeram sentido em outros casos, e as frases que pareciam corretas devido a um conjunto de circunstâncias pareciam erradas nos outros – muitas vezes sutilmente.*”⁸⁶ (KLEIN, 2013)

⁸⁶ Texto original: “Editing one narrative does not mean you’ve edited them all. (...) edits that made sense in one case ended up not working in other cases, and sentences that seemed correct given one set of circumstances seemed wrong in others – often subtly.” (KLEIN, 2013)

Figura 11 – Notícia automatizada publicada pela ProPublica com o perfil de uma das 52 mil escolas que fazem parte do projeto *The Opportunity Gap*



Fonte: Reprodução de internet⁸⁷

⁸⁷ Disponível em: <http://projects.propublica.org/schools/schools/201226001476> Acessado em: 08/10/2017

Apesar das experiências com outros projetos da *Narrative Science*, Kristian Hammond também reconheceu o desafio do projeto desenvolvido com a *ProPublica* de lidar com um banco de dados extenso e a produção concomitante de 52 mil notícias, o que acabou gerando alguns erros. *“Mesmo que os dados se pareçam os mesmos, há tantas condições ambientais diferentes que estão fora do escopo desses dados que as comparações não funcionaram e, de fato, estavam criando falsas analogias.”*⁸⁸ (LECOMPTE, 2015)

Figura 12 – Arquivos em formato JSON das notícias automatizadas produzidas pela Narrative Science para o projeto *The Opportunity Gap* da *ProPublica*

school-report-010114000411.json	Jan 4, 2013 1:35 PM	2 KB	JSON
school-report-010114001860.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-010207000790.json	Jan 4, 2013 1:34 PM	2 KB	JSON
school-report-010222000846.json	Jan 4, 2013 1:34 PM	2 KB	JSON
school-report-010265001573.json	Jan 4, 2013 1:35 PM	2 KB	JSON
school-report-010297000152.json	Jan 4, 2013 1:31 PM	2 KB	JSON
school-report-040724002267.json	Jan 4, 2013 1:32 PM	2 KB	JSON
school-report-040906000939.json	Jan 4, 2013 1:35 PM	2 KB	JSON
school-report-060001409082.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-060681000608.json	Jan 4, 2013 1:32 PM	2 KB	JSON
school-report-060744002334.json	Jan 4, 2013 1:32 PM	2 KB	JSON
school-report-060846011326.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-061350001542.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-061488001858.json	Jan 4, 2013 1:34 PM	2 KB	JSON
school-report-061518001908.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-062019002414.json	Jan 4, 2013 1:32 PM	2 KB	JSON
school-report-062271002889.json	Jan 4, 2013 1:32 PM	2 KB	JSON
school-report-062271002954.json	Jan 4, 2013 1:33 PM	2 KB	JSON
school-report-062361003588.json	Jan 4, 2013 1:34 PM	1 KB	JSON
school-report-062547005903.json	Jan 4, 2013 1:35 PM	2 KB	JSON
school-report-062628003923.json	Jan 4, 2013 1:34 PM	2 KB	JSON
school-report-062769004160.json	Jan 4, 2013 1:33 PM	2 KB	JSON

Fonte: KLEIN, 2013

Ao final, Klein recebeu as 52 mil notícias em arquivos com a extensão JSON via FTP (Figura 12) que foram importadas para o aplicativo de publicação de notícias da *ProPublica* e, de um modo geral, ele avalia positivamente o uso de *softwares* NLG nas redações: *“A geração algorítmica de histórias parece ser uma solução intrigante para o problema de escalonamento de narrativas no jornalismo na era do big data. Nos ajudou a adaptar o nosso jornalismo de dados para uma nova audiência, e a criar histórias em escalas, o que seria impensável de outra forma.”*⁸⁹ (KLEIN, 2013)

⁸⁸ Texto original: “Even though the data look the same, there are so many different environmental conditions that are outside the scope of this data that the comparisons would not fly and would, in fact, be making false analogies.” (LECOMPTE, 2015)

⁸⁹ Texto original: “algorithmic story generation seems like an intriguing solution to the problem of scaling narrative journalism in the era of big data. It’s helped us tune our data journalism for a new audience, and helped create stories at scale that would have been unthinkable otherwise.” (KLEIN, 2013)

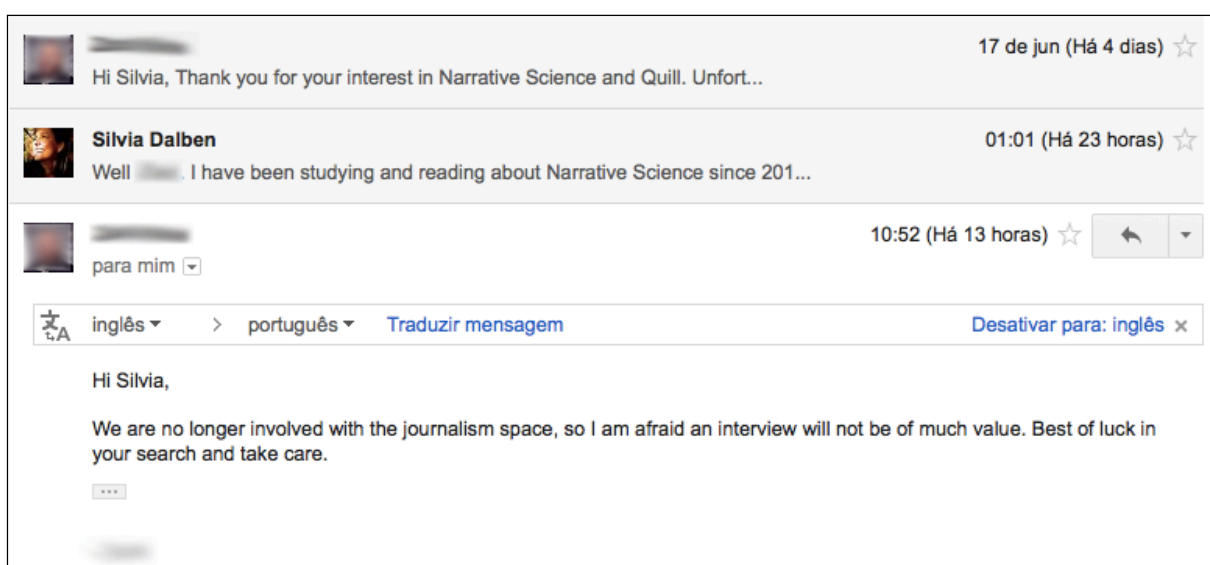
4.1.5 Mudança de rumo e abandono do jornalismo

Ainda em 2012, quando anunciou a parceria com a *Forbes*, a *Narrative Science* já vislumbrava atrair novos mercados associados, principalmente, a sistemas de *Business Intelligence* (BI). David Rosenblatt, um dos executivos da *startup*, explica em uma reportagem para a *Wired* na época que o mercado da *Narrative Science* era “transformar números em palavras” e que “jornalismo parece ser apenas um ruído – o filé devem ser os relatórios de gerenciamento.”⁹⁰ (LEVY, 2012)

E foi seguindo esta premissa que a companhia foi direcionando a sua estratégia. Em junho de 2013, a *Narrative Science* recebeu um aporte de investimento do fundo *In-Q-Tel* vinculado à *CIA*, que foi definido como uma “parceria estratégica” pelo executivo Steve Bowsher. “A plataforma de Inteligência Artificial da *Narrative Science* analisa dados e comunica essa informação de maneira fácil de ler e de entender, o que será de grande valia para nossos clientes na comunidade da inteligência.” (SÁ, 2013)

Além da parceria com a *CIA*, a *Narrative Science* começou a se associar a empresas para a geração automatizada de relatórios gerenciais, como o *Credit Suisse*, a *Mastercard*, a *T. Rowe Price*, a *USAA* e a *American Century Investments Services Inc.* Em 2015, a *startup* foi listada pela *CNBC* como uma das 50 empresas mais disruptivas do mundo, ao lado de *startups* como *Airbnb*, *Pinterest*, *Spotify* e *Snapchat*.

Figura 13 – Email da *Narrative Science* confirma a mudança de rumo da empresa



Fonte: Email pessoal da pesquisadora

⁹⁰ Texto original: “journalism might be only the sizzle—the steak might be management reports.” (LEVY, 2012)

Em outubro deste mesmo ano, a *Forbes* publicou a última notícia em parceria com a *Narrative Science* e, ao que tudo indica, a empresa de Chicago parou de se dedicar ao Jornalismo Automatizado. Em busca de mais informações, de acesso ao demo do *software Quill* e/ou de uma possível entrevista por *Skype*, eu entrei em contato por email com a *Narrative Science* em junho de 2016, e obtive como resposta que eles não atuavam mais no mercado jornalístico, como é possível ver na Figura 13.

4.2 Automated Insights: StatSheet e a parceria com a Associated Press

4.2.1 StatSheet

Em 2010, alguns meses após o lançamento da *Narrative Science*, uma outra *startup* começou a chamar a atenção com uma proposta bem parecida. Após alavancar um investimento de 1.3 milhões de dólares em agosto de 2010, o ex-engenheiro da *Cisco* Robbie Allen contratou um time de programadores e lançou a *startup StatSheet*. Sua intenção, como Allen mesmo descreve, era de que ela se tornasse “*um novo tipo de empresa de mídia esportiva focada em tecnologia e automação, e não em pagar pessoas para pesquisar e reportar como na mídia esportiva tradicional.*”⁹¹ (TSOTSIS, 2010)

Três meses depois e com uma equipe de apenas nove pessoas, a *StatSheet* anunciou o lançamento de uma rede com 345 sites - um para cada time de basquete da primeira divisão da liga universitária dos Estados Unidos – com a publicação média de 15 mil notícias por mês, todas escritas por *software*. (DALEN, 2012)

Na figura 14, é possível observar uma notícia automatizada publicada no site dedicado ao time de basquete *Ohio State* da *StatSheet network*. Na época, Robbie Allen afirmava que “*o único envolvimento humano era com a criação do algoritmo que gera as postagens*”⁹² (SCHONFELD, 2010), e tinha uma postura bem crítica quanto ao trabalho realizado por repórteres esportivos, como frisou em uma entrevista. “*Alguém me enviou um artigo escrito por um colunista esportivo bem conhecido, e o nosso algoritmo poderia ter feito aquilo.* (...)”

⁹¹ Texto original: “a new type of sports media company that is focused on technology and automation instead of paying people to research and report news like is the norm with traditional sports media.” (TSOTSIS, 2010)

⁹² Texto original: “The only human involvement is with creating the algorithm that generate the posts” (SCHONFELD, 2010)

Escrever conteúdo esportivo é muito tedioso. Você está cobrindo o desempenho, as estatísticas e os jogadores.”⁹³ (GRIMM, 2010)

Figura 14 – Notícia automatizada produzida pela StatSheet

The screenshot shows the 'BUCKEYES BEAT' website, which is powered by StatSheet. The header includes navigation links for 'STATSHEET NETWORK', 'NCAA BASKETBALL', 'NFL', and 'MLB'. A search bar is present with filters for 'FIND ARTICLES:', 'CHANNEL', 'CATEGORY', and 'GAME'. The main content area features a large article titled 'Jared Sullinger Wins the StatSheet NCAA Freshman of the Week - January 31st, 2011'. The article text describes Sullinger's performance, mentioning his 37.5 points per game and 5.5 three-pointers per game. It also mentions other players like Jimmer Fredette and Nolan Smith. The article is categorized under 'NCAA POTW' and includes a 'Share this article' section with a 'Curtir' button. To the right of the article, there are sections for 'NCAA, NIT BRACKETS', 'Fan Satisfaction', and 'Categories'. The 'Fan Satisfaction' section includes a 'Find another NCAA team:' search bar. The 'Categories' section lists various links like 'Boxscore', 'Game Preview', 'General', 'Injury Report', 'On this Day', 'Season Preview', 'Top 25 Update', 'Conf POTW', 'Game Recap', 'Highlights', 'NCAA POTW', 'Recap Notes', and 'StatSeed Update'. At the bottom of the article, there are 'Weekly Leaders' and 'Tags'.

STATSHEET NETWORK NCAA BASKETBALL CHANGE TEAM ▼ NFL MLB FIND ARTICLES: CHANNEL ▼ CATEGORY ▼ GAME ▼

BUCKEYES BEAT
Ohio State basketball stats and analysis powered by StatSheet

Follow Buckeyes Beat
f t g a r More →
Search Buckeyes Beat Search

HOME NEWS SCHEDULE STATS ROSTER COMPARE VIDEO ABOUT SHOP

TOP STATSHEET APPS
STATSMACK / <EMBED>

WSU 42 @ 73 WIN OSU
Fri, Nov 11, 9:00PM ET
Recap | Boxscore | Scoreboard

UF 74 @ 81 WIN OSU
Tue, Nov 15, 8:00PM ET
Recap | Boxscore | Scoreboard

JSU 41 @ 85 WIN OSU
Fri, Nov 18, 9:00PM ET
Recap | Boxscore | Scoreboard

Jared Sullinger Wins the StatSheet NCAA Freshman of the Week - January 31st, 2011

Filed under **NCAA POTW** on January 31st, 2011

Share this article
or Curtir Cadastre-se para ver do que seus amigos gostam.

After a quality outing which included facing the #4 team in the nation, Jimmer Fredette, a 6-2 senior from Brigham Young, has been selected StatSheet NCAA Player of the Week. StatSheet's NCAA Freshman of the Week is Jared Sullinger and Utah State's Brian Green is StatSheet's NCAA 6th Man of the Week.

Fredette put up a strong 37.5 points per game and 5.5 three pointers per game in a victory against 4th ranked San Diego State and a loss to New Mexico.

Fredette led all scoring for the Cougars with 43 points on 58% shooting. in a 71-58 upset win against San Diego State. He then led the team in scoring (32) and added an additional 7 assists in an 86-77 loss to New Mexico.

This is the first StatSheet NCAA Player of the Week award for Jimmer Fredette. The runner-up this week is senior Nolan Smith from Duke. Honorable mention goes to Notre Dame senior Ben Hansbrough.

StatSheet NCAA Freshman of the Week: Ohio State's Jared Sullinger

Jared Sullinger played an average of 32 minutes starting both games for the Buckeyes. In a mammoth 87-64 win against Purdue, Sullinger churned out 17 points and 7 rebounds. Against Northwestern, in a tight 58-57 win, he led scoring and rebounding with 21 and 8.

This is a first for Sullinger, who has never won the StatSheet NCAA Freshman of the Week award. The runner-up is Jeremy Lamb from Connecticut. Honorable mention goes to Doron Lamb from Kentucky.

StatSheet NCAA 6th Man of the Week: Utah State's Brian Green

Senior Brian Green averaged 34 minutes over two games for the Aggies. In an 84-65 win over San Jose State, Green led the Aggies scoring and rebounding with 25 and 6. Then while escaping Hawaii, 89-84, he led all scoring with 22 points and added 3 assists.

This is the first time this season Green has won the StatSheet NCAA 6th Man of the Week award. Matt Kenney, a sophomore guard from Valparaiso, is runner-up. Faisal Aden, a junior guard from Washington State, wins honorable mention.

Weekly Leaders

- Scoring: Steven Pledger, Oklahoma: 38 PPG
- Rebounding: Kenneth Faried, Morehead State: 18 RPG
- Assists: Jerrell Thompson, Sacred Heart: 10 APG
- Blocked Shots: William Mosley, Northwestern State: 6.5 BPG
- Free Throws: Xavier Silas, Northern Illinois: 14 PPG
- Three Pointers: Steven Pledger, Oklahoma: 7 TPG
- Steals: Troy Payne, Santa Clara: 6 SPG

Tags: Jon Diebler, Jared Sullinger

NCAA, NIT BRACKETS

Fan Satisfaction
Install or upgrade your Flash Player to view StatSheet charts.
[More about Fan Satisfaction](#)

Find another NCAA team:

Categories

- Boxscore
- Game Preview
- General
- Injury Report
- On this Day
- Season Preview
- Top 25 Update
- Conf POTW
- Game Recap
- Highlights
- NCAA POTW
- Recap Notes
- StatSeed Update

Sign up to get updates in your inbox:
 GET UPDATES!

Fonte: Reprodução de internet⁹⁴

⁹³ Texto original: "Someone sent me an article written by a well-known sports columnist, and our algorithm could have generated that (...) Writing sports content is still really tedious. You're covering a performance, stats and players." (GRIMM, 2010)

⁹⁴ Disponível em: <https://web.archive.org/web/20120830032017/http://buckeyesbeat.com/ohio-state-basketball/ncaa-potw/jared-sullinger-wins-statsheet-ncaa-freshman-week-january-31st-2011> Acessado em 08/09/2017

Por outro lado, na mesma entrevista, Allen também assume as limitações do conteúdo produzido por seus algoritmos: *“O conteúdo pode ser um pouco duro em algumas partes, mas você não tem muitos vencedores do Pulitzer escrevendo sobre esportes, então o nível não é muito alto.”*⁹⁵ (GRIMM, 2010) Apesar de polêmico, Allen não acredita que esta inovação irá retirar o emprego de ninguém. *“Eu não acho que vou tirar o emprego de ninguém. Se algo acontecer, esta tecnologia vai liberar os repórteres para se concentrarem nos eventos mais interessantes do esporte ao invés de ficarem sentados fazendo um monte de cálculos numéricos.”*⁹⁶ (Idem)

Em resposta à polêmica em torno do lançamento da *StatSheet*, o jornalista Jeff Bercovici, da *Forbes* levanta uma importante questão ética em torno da autoria das notícias escritas por algoritmos, destacando que os 345 sites da *StatSheet Network* não revelam aos leitores que aqueles conteúdos não foram escritos por jornalistas. Ele destaca que o site só afirmava que a *StatSheet* *“utiliza tecnologia avançada para garantir que você está tendo acesso à análise mais precisa possível.”*⁹⁷ (BERCOVICI, 2010) Bercovici ainda questiona um dos objetivos de Allen, revelado na época para o *The New York Times*: *“Meu objetivo era que 80 por cento dos leitores não questionassem se o conteúdo foi escrito por um humano, e agora que lançamos, eu acredito que este percentual é ainda maior.”*⁹⁸ (STROSS, 2010)

4.2.2 Uma parceria com a *Associated Press*

Em 2011, a *StatSheet* mudou seu nome para *Automated Insights* e anunciou a expansão de sua tecnologia para a cobertura de finanças e imóveis⁹⁹, mas a empresa só voltou a ganhar destaque em 2014, quando anunciou uma parceria com a agência de notícias *Associated Press* (AP). A demanda partiu da própria AP, que no verão de 2013 decidiu, em uma reunião com executivos, experimentar o uso de Inteligência Artificial para a produção automatizada de notícias, com foco no mercado financeiro e nos esportes. (MARCONI, SIEGMAN, 2017)

⁹⁵ Texto original: “The content may be a little rough in some places, but you don’t have many Pulitzer winners writing sports, so the bar isn’t very high” (GRIMM, 2010)

⁹⁶ Texto original: “I’m not out to put anybody job and I don’t think I will. If anything, it’s going to free up writers to focus on the more interesting events of sports instead of having to sit down and do a bunch of number crunching.” (Idem)

⁹⁷ Texto original: “utilizes advanced technology to ensure that you are getting the most accurate analysis possible.” (BERCOVICI, 2010)

⁹⁸ Texto original: “My goal was that 80 percent of readers wouldn’t question that the content was written by a human, and now that we’ve launched, I think the percentage is higher” (STROSS, 2010)

⁹⁹ Informação retirada do verbete da *Wikipedia* sobre a *StatSheet*. Disponível em: <https://en.wikipedia.org/wiki/StatSheet> Acessado em: 08/09/2017.

As primeiras notícias automatizadas começaram a ser publicadas em 2014 (Figura 15) e, de acordo com a editora Lisa Gibbs, a AP passaria a publicar com esta tecnologia doze vezes mais notícias durante uma temporada de resultados financeiros, incluindo pequenas empresas que nunca receberam muita atenção. (Ibidem) A estimativa era de que, com o uso do *software*, os jornalistas teriam 20% do seu tempo liberado para realizarem outros trabalhos mais complexos e análises qualitativas dos resultados. *"Com o tempo liberado, os jornalistas da AP podem se engajar mais na produção de conteúdos gerados pelos usuários, desenvolver relatórios multimídia, aprofundarem uma pesquisa e focarem em notícias mais complexas."*¹⁰⁰ (Ibidem)

Para a *Automated Insights* implantar o *Wordsmith* na redação da *Associated Press*, a editora-assistente Philana Patterson trabalhou com programadores que traduziram as regras e os modelos de notícias sobre balancetes financeiros em código algorítmico, o que possibilitou a geração de textos automatizados a partir do processamento de dados disponibilizados pela *Zack Investment Research* (LECOMPTE, 2015), a mesma fonte utilizada na época pela *Forbes* e *Narrative Science*. Para Patterson, um dos desafios deste projeto foi pensar no maior número de variáveis possível para repassar as situações aos programadores, e *"mesmo assim você não consegue pensar em todas as variáveis"*¹⁰¹. (LECOMPTE, 2015)

Patterson afirma que as notícias automatizadas são publicadas entre seis e dez minutos após a divulgação dos dados pela *Zacks Investment Research*, e que erros são muito raros e só ocorrem quando o *input* de dados está incorreto (AP ARCHIVE, 2015), informação que é corroborada pela editora Lisa Gibbs: *"algoritmos não cometem erros de digitação ou de cálculos aritméticos. Os erros geralmente ocorrem por um problema com os dados. Se os dados estiverem ruins, você terá uma notícia ruim."*¹⁰² (MARCONI, SIEGMAN, 2017)

Como mostra a figura 15, todas as notícias automatizadas publicadas pela *Associated Press* são assinadas com o seguinte texto: *"Esta notícia foi gerada pela Automated Insights com dados da Zacks Investment Research"*¹⁰³, mas nem todos os clientes da *Automated Insights* revelam que suas notícias foram escritas por um "robô", pois temem que isto chame a atenção dos leitores que terão uma postura mais crítica quanto aos textos, procurando

¹⁰⁰ Texto original: "With the freed-up time, AP journalists are able to engage with more user-generated content, develop multimedia reports, pursue investigative work and focus on more complex stories." (MARCONI, SIEGMAN, 2017)

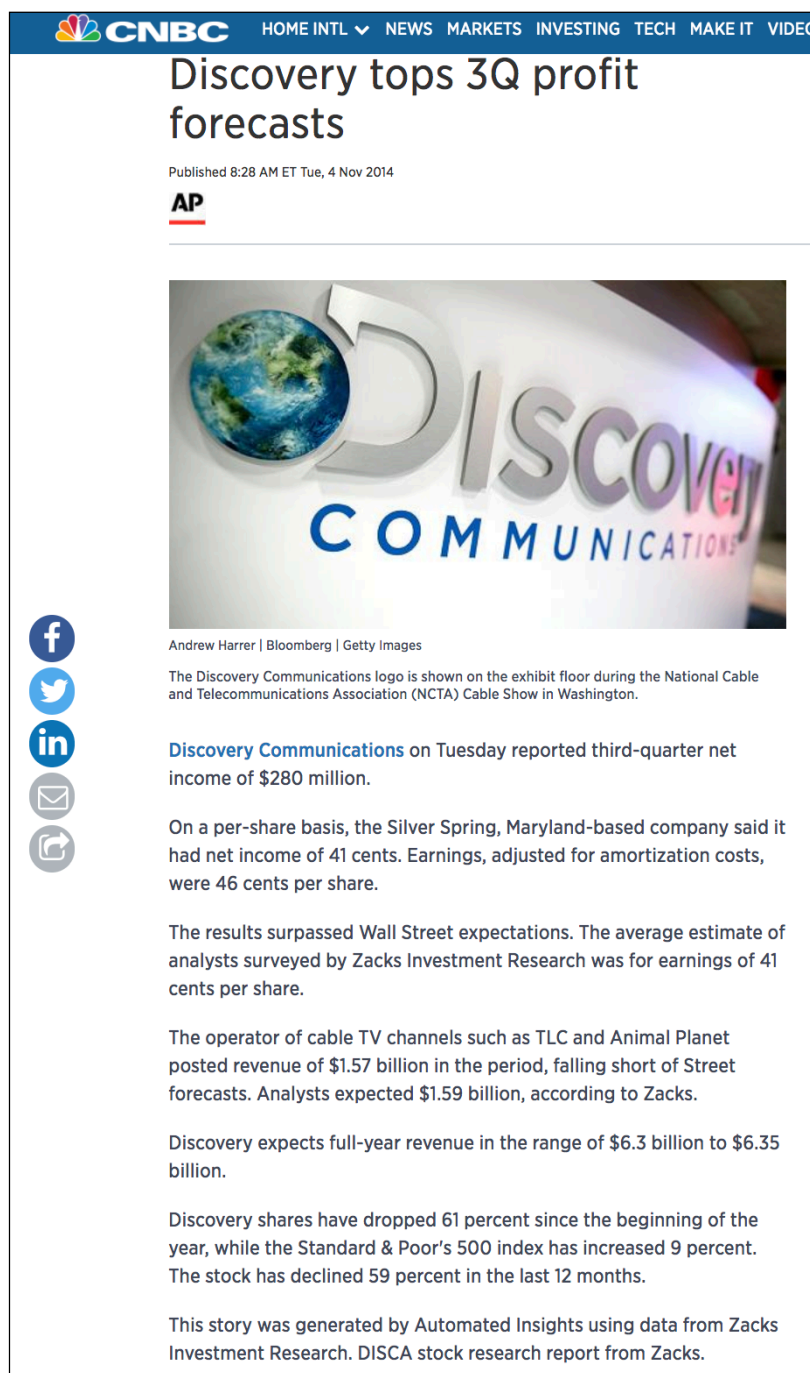
¹⁰¹ Texto original: "You have to think of as many variables as you can, and even then you might not think of every variable," (LECOMPTE, 2015)

¹⁰² Texto original: "algorithms don't make typos or arithmetic miscalculations. The errors are generally because of a problem with the data. If the data's bad, you get a bad story." (MARCONI, SIEGMAN, 2017)

¹⁰³ Texto original: "This story was generated by Automated Insights using data from Zacks Investment Research".

defeitos por ele ter sido escrito por um algoritmo. Um outro ponto levantado seria um tratamento diferenciado pelos algoritmos de ranqueamento de sistemas de busca, como o *Google*, que poderiam dar um tratamento diferenciado para as notícias automatizadas e qualificá-las de forma inferior. (KIRKLAND, 2014)

Figura 15 – Notícia automatizada publicada pela Associated Press



Discovery tops 3Q profit forecasts

Published 8:28 AM ET Tue, 4 Nov 2014

AP

Andrew Harrer | Bloomberg | Getty Images

The Discovery Communications logo is shown on the exhibit floor during the National Cable and Telecommunications Association (NCTA) Cable Show in Washington.

Discovery Communications on Tuesday reported third-quarter net income of \$280 million.

On a per-share basis, the Silver Spring, Maryland-based company said it had net income of 41 cents. Earnings, adjusted for amortization costs, were 46 cents per share.

The results surpassed Wall Street expectations. The average estimate of analysts surveyed by Zacks Investment Research was for earnings of 41 cents per share.

The operator of cable TV channels such as TLC and Animal Planet posted revenue of \$1.57 billion in the period, falling short of Street forecasts. Analysts expected \$1.59 billion, according to Zacks.

Discovery expects full-year revenue in the range of \$6.3 billion to \$6.35 billion.

Discovery shares have dropped 61 percent since the beginning of the year, while the Standard & Poor's 500 index has increased 9 percent. The stock has declined 59 percent in the last 12 months.

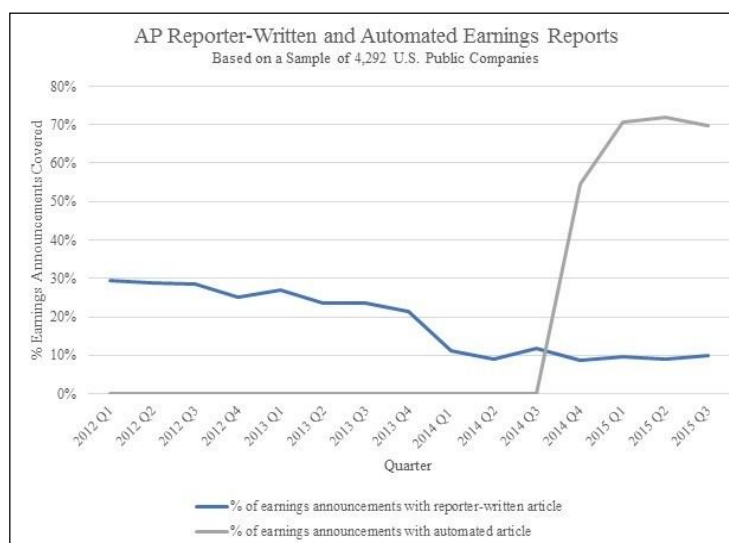
This story was generated by Automated Insights using data from Zacks Investment Research. DISCA stock research report from Zacks.

Fonte: Reprodução de internet do site da CNBC¹⁰⁴

¹⁰⁴ Disponível em: <https://www.cnbc.com/2014/11/04/discovery-tops-3q-profit-forecasts.html> Acessado em: 23/10/2017

Em uma análise do impacto da publicação de notícias automatizadas no mercado financeiro, uma pesquisa realizada pela *Stanford University* e pela *University of Washington* concluiu que estes conteúdos ajudaram a “aumentar o volume de negociação e a liquidez das empresas”¹⁰⁵. (MARCONI, 2016) Uma das principais contribuições desta pesquisa, como mostra a Figura 16, é um gráfico que compara o número de relatórios financeiros publicados pela *Associated Press* escritos por repórteres e por *software* entre 2012 e 2015. Este gráfico foi produzido a partir da análise de dados de 4.292 empresas e 57.467 artigos publicados pela *Associated Press*, repercutindo os resultados financeiros trimestrais de empresas listadas na bolsa. (Idem)

Figura 16 – Gráfico compara notícias sobre o mercado financeiro publicadas pela Associated Press escritas por repórteres e por software entre 2012 e 2015



Fonte: MARCONI, 2016¹⁰⁶

Um dos *insights* que a leitura deste gráfico nos mostra é a de que houve uma queda abrupta nas publicações de balancetes financeiros escritos por jornalistas a partir do primeiro trimestre de 2014, seguida pelo início da publicação dos balancetes automatizados produzidos pelo *software* da *Automated Insights* a partir do terceiro semestre do mesmo ano e com uma cobertura sete vezes maior. A *Associated Press* continuou publicando balancetes financeiros escritos por jornalistas após a implementação do *software*, mas o volume de publicações caiu em um terço se compararmos o primeiro trimestre de 2012 e a média a partir do último trimestre de 2014.

¹⁰⁵ Texto original: "The analysis conducted found compelling evidence that these articles increase firms' trading volume and liquidity." (MARCONI, 2017)

¹⁰⁶ Disponível em: <https://insights.ap.org/industry-trends/study-news-automation-by-ap-increases-trading-in-financial-markets> Acessado em: 18/10/2017

O aumento no volume de notícias produzidas pela *Associated Press* com a automatização é exaltado por Joe Procopio, gerente de produto da *Automated Insights*. “Antes do *Wordsmith*, a *Associated Press* escrevia sobre cerca de 440 empresas em cada temporada de resultados. Depois do *Wordsmith*, eles podem escrever quase cinco mil artigos por temporada e este número está crescendo.”¹⁰⁷ (AL JAZEERA, 2015)

Para Robbie Allen, CEO da *Automated Insights*, o Jornalismo Automatizado segue uma outra lógica e, ao invés de imaginarmos as notícias sendo lidas por milhares de pessoas como desejam os jornalistas, o objetivo da empresa é exatamente o oposto: criar um milhão de artigos para serem lidos, cada um, por uma pessoa. “Não acho que os jornalistas devam ficar preocupados pois estamos criando conteúdo que não existia antes.”¹⁰⁸ (KIRKLAND, 2014)

Lou Ferrara, vice-presidente da *Associated Press*, destaca ainda que não houve nenhuma redução no número de empregos (AL JAZEERA, 2015) e que o uso desta tecnologia permitiu liberar os jornalistas de tarefas repetitivas, rotineiras e entediadas que podem ser facilmente reproduzidas por uma máquina com a mínima supervisão humana. “Isto nos permitirá empregar nossos recursos humanos de maneira mais criativa durante a temporada de resultados. Assim descobriremos tendências e informações exclusivas para publicar ao mesmo tempo que esses dados.” (PEREDA, 2014)

A editora-assistente Philana Patterson concorda com Ferrara quanto ao objetivo de liberar os jornalistas de tarefas repetitivas, e cita um exemplo real de como isto acontece na prática.

Nós definitivamente estamos tendo mais tempo para trabalhar com notícias factuais, trabalhos investigativos e fazendo notícias mais matizadas, textos mais detalhados. Um exemplo que eu realmente gostei neste trimestre foi com os resultados da *Apple*. A gente publicou a notícia automatizada que realmente só tinha o básico e daí o nosso repórter Brandon Bailey pode se dedicar a uma história mais detalhada sobre as vendas de *iPhones*. Então, fomos capazes de oferecer o básico e, em seguida, uma notícia mais detalhada sobre o que todo mundo queria saber, que eram as vendas de *iPhones*.¹⁰⁹ (AP ARCHIVE, 2015)

¹⁰⁷ Texto original: “Before *Wordsmith*, The *Associated Press* would write about around four hundred and forty companies per earning season. After *Wordsmith*, they can write close to five thousand articles on these companies per season and that number is growing” (AL JAZEERA, 2015)

¹⁰⁸ Texto original: “I don't think journalists have anything to worry about because we're creating content where it didn't exist before” (KIRKLAND, 2014)

¹⁰⁹ Texto original: “We are definitely able to spend more time working on breaking news, investigative work and doing more nuanced stories, things that go into more detail. You know, one example that I really like this quarter was with *Apple* earnings. We had the automated story out there which really got just the basics out and then our reporter Brandon Bailey was able to go into a more nuanced detailed story on *iPhone* sales. So we were able to offer members the basics and then a more detailed story about what everybody wants to know about, which is the sale of *iPhones*.” (AP ARCHIVE, 2015)

Em 2015, a *Associated Press* contratou seu primeiro “editor de notícias automatizadas”, Justin Myers, cuja função é identificar tarefas repetitivas e onerosas desempenhadas pela equipe de jornalistas que podem ser automatizadas. “*Vamos deixar os computadores fazerem aquilo que eles são bons em fazer, e deixar os jornalistas humanos fazendo aquilo que só eles sabem fazer.*”¹¹⁰ (LECOMPTE, 2015) Em uma entrevista para o relatório da *Associated Press* sobre “*O futuro das redações na era das máquinas inteligentes*”¹¹¹, Justin Myers explica como produzir conteúdos automatizados com qualidade:

Qualidade depende principalmente de duas coisas: explicação e teste. Os editores devem ser capazes de explicar o que precisam que estes sistemas façam, bem como apontar problemas que eles detectam ou preveem. Os técnicos de implementação devem ser capazes de explicar como seus sistemas produziram resultados específicos, tanto para fins de transparência editorial quanto para resolução de problemas. E ambos grupos precisam se sentir confortáveis, testando frequentemente estes sistemas – e suas expectativas – antes e depois do lançamento.¹¹² (MARCONI, SIEGMAN, 2017)

Existe uma visão na *Associated Press* de que o uso de Inteligência Artificial aplicado ao jornalismo seria o resultado de um trabalho em equipe envolvendo jornalistas, programadores e cientistas de dados, o que abriria oportunidades para a atuação de profissionais de outras áreas. Um exemplo disso é o matemático Larry Fenn que trabalha no time de jornalismo de dados da agência. Fenn ressalta que “*é importante trazer a ciência para as redações porque os padrões de uma boa ciência – transparência e reprodutibilidade – se encaixam perfeitamente no jornalismo*”¹¹³. (MARCONI, SIEGMAN, 2017)

4.3 Do jornalismo à automatização: a experiência do *Los Angeles Times*

4.3.1 Uma nova forma de reportar homicídios

Em 2010, uma equipe de jornalistas do *Los Angeles Times* experimentava uma outra forma de automatizar a produção de conteúdo para seu site, bem diferente do que já havia

¹¹⁰ Texto original: “Let’s have a computer do what a computer’s good at, and let’s have a human do what a human’s good at” (LeCompte, 2015)

¹¹¹ Texto original: *A guide for newsrooms in the age of smart machines* Disponível em: https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf

¹¹² Texto original: “Quality depends on two main things: explanation and testing. Editorial stakeholders must be able to explain what they need these systems to do, as well as flag problems they observe or expect. Technical implementers must be able to explain how their systems produced particular results, both to support editorial transparency and to resolve errors. And both groups need to be comfortable frequently testing these systems — and their assumptions — before and after launch.” (MARCONI, SIEGMAN, 2017)

¹¹³ Texto original: “It’s important to bring science into newsrooms because the standards of good science — transparency and reproducibility — fit right at home in journalism,” (MARCONI, SIEGMAN, 2017)

sido criado para a redação de previsões do tempo e de notícias sobre jogos universitários. O tema da vez eram os homicídios.

O *The Homicide Report* [homicide.latimes.com] era um projeto antigo do jornal, lançado em 2007 em formato de blog e com a desafiadora missão de escrever uma notícia para cada homicídio que acontecesse na região metropolitana de Los Angeles, uma área que abrange 88 municípios. A média de ocorrências na época era de mil casos por ano, e a responsável por escrever estas notícias era a jornalista Jill Leovy. A premissa do site, desde o início, era buscar uma forma igualitária para tratar cada um dos homicídios registrados na região, em um esforço de ir contra o “valor notícia” que acaba dando mais atenção para casos envolvendo uma adolescente branca ou um assassinato em massa, e não dá visibilidade para casos mais frequentes envolvendo jovens negros de 17 a 22 anos mortos nas regiões mais pobres da cidade. (REID, 2014)

Em uma entrevista na época do lançamento do site para o programa *On The Media* da rádio WNYC, Leovy explica o objetivo do projeto.

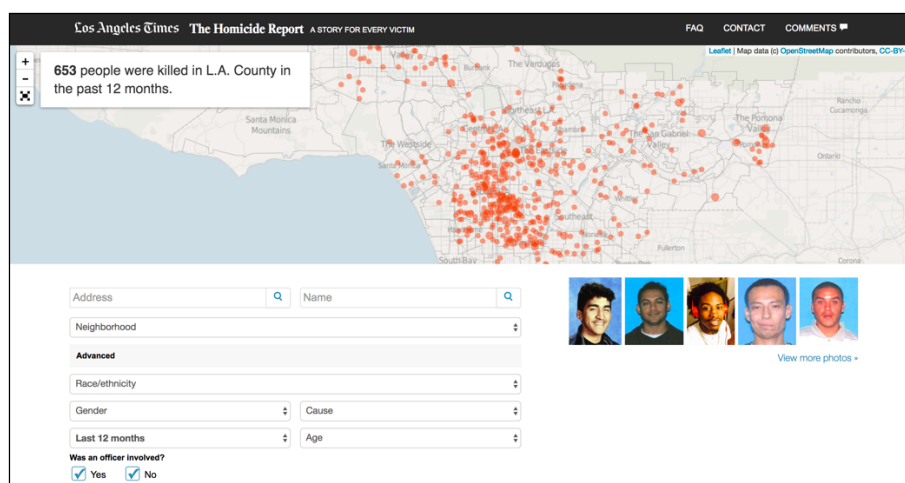
Há uma espécie de lógica invertida na cobertura de homicídios pela imprensa, onde a natureza da notícia é cobrir acontecimentos *man-bites-dog* (poucos usuais), e o que você acaba fazendo é cobrir as estatísticas de homicídios. Você cobre casos muito improváveis que não representam o que realmente acontece na maioria dos dias na região de Los Angeles. Isso cria, penso eu, uma visão falsa de quem está seguro e quem não está e onde o homicídio está concentrado. E o *The Times* queria dar uma visão precisa de quais homicídio acontecem, tanto longe quanto perto.¹¹⁴ (ON THE MEDIA, 2007)

Leovy ficou a frente do blog por aproximadamente um ano, mas encontrou dificuldades em levar o projeto adiante por questões técnicas e de dados disponíveis que viabilizassem a cobertura dos casos. Por um tempo, o blog foi mantido por jornalistas recém-formados, e parou de ser atualizado em novembro de 2008. (YOUNG, HERMIDA, 2014)

Quatorze meses depois, o *Homicide Report* foi relançado em janeiro de 2010 com uma nova plataforma que incluía uma ferramenta de geolocalização indicando o local exato do crime e um sistema de busca que permitia aos leitores navegar pelos conteúdos utilizando filtros como raça, gênero, causa da morte, envolvimento ou não da polícia, bairro, ano, como mostra a Figura 17.

¹¹⁴ Texto original: “There’s sort of an upside-down logic of press coverage in homicide where the nature of news is to cover the man-bites-dog story, and what you end up doing is you cover the statistical fringe of homicide. You cover the very unlikely cases that don’t represent what’s really happening most days in Los Angeles County. It creates, I think, a false view of who’s safe and who’s not and where homicide is concentrated. And The Times wanted to give an accurate view of what homicide looks like, both from afar and close up” (ON THE MEDIA, 2007)

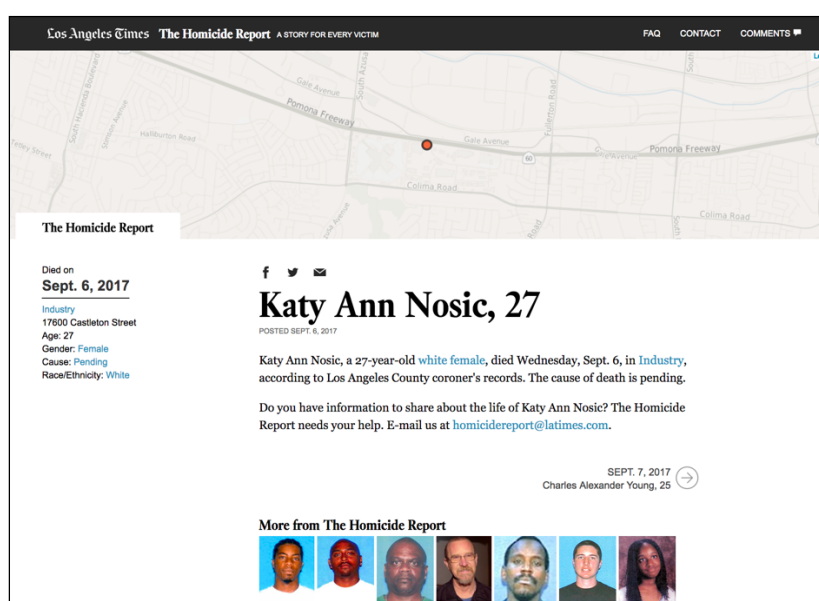
Figura 17 – Capa do *Homicide Report* no site do *Los Angeles Times*



Fonte: Reprodução de internet¹¹⁵

No relançamento do *Homicide Report*, a principal novidade foi a implementação de um sistema que escrevia e publicava automaticamente pequenas notícias sobre os homicídios. O *robô-post* (YOUNG, HERMIDA, 2014), como foi denominado pelos jornalistas envolvidos no projeto, era publicado automaticamente no site para cada uma das vítimas registradas no banco de dados fornecido pelo *Los Angeles County coroner's office*. O texto era breve e continha informações como data, localização do homicídio, raça, nome, idade da vítima e em qual bairro ocorreu o crime, como é possível observar na Figura 18.

Figura 18 – Post automatizado publicado no *Homicide Report*



Fonte: Reprodução de internet¹¹⁶

¹¹⁵ Disponível em: <http://homicide.latimes.com> Acessado em 08/09/2017

¹¹⁶ Disponível em: <http://homicide.latimes.com/post/katy-ann-nasic/> Acessado em 08/09/2017

Um dos responsáveis pelo desenvolvimento do site, o jornalista e programador Ken Schwencke, afirmou em uma entrevista a pesquisadores da *University of British Columbia* que o algoritmo responsável pela atualização automática do *Homicide Report* é extremamente simples.

É um tipo simples de inovação, um tipo simples de ferramenta que tem um forte impacto. Faz muito com muito pouco. Não é um pedaço complicado de código, não é uma teoria complicada, não é um algoritmo complicado. É uma coisa simples que construímos que tem um impacto desproporcional em relação à quantidade de tempo que efetivamente gastamos nela.¹¹⁷ (YOUNG, HERMIDA, 2014)

A partir do *robô-post* inicial, os repórteres decidem quais homicídios serão atualizados com mais detalhes, e a apuração dos casos se torna ainda mais completa a partir de informações fornecidas por e-mails e comentários dos leitores.

Figura 19 – Notícia atualizada por jornalista no *Homicide Report*

Los Angeles Times **The Homicide Report** A STORY FOR EVERY VICTIM

FAQ CONTACT COMMENTS

Todd James Britt, 51

By **NICOLE SANTA CRUZ**

POSTED SEPT. 23, 2015

Todd James Britt, a 51-year-old black man, was shot and killed Wednesday, Sept. 16, in the 7900 block of South Normandie Avenue in Vermont Knolls, according to Los Angeles County coroner's records.

Shortly before 6 p.m., Britt was walking his dog north on South Normandie Avenue when he was approached by a man who shot him, said Los Angeles Police Department Det. Dave Garrido. It's unclear whether the man said anything before he shot Britt.

The gunman then ran to a waiting vehicle that fled east on West 79th Street, according to preliminary reports.

Britt was taken to a hospital where he was pronounced dead at 7:23 p.m., according to coroner's records. The dog was uninjured in the shooting.

Anyone with information is asked to call the Criminal Gang Homicide Division at (323) 786-5113. Those who wish to remain anonymous can call Crime Stoppers at (800) 222-8477.

Contact the Homicide Report. Follow @nicosantacruz and @latimeshomicide on Twitter.

The Homicide Report

Todd James Britt, 51

Died on **Sept. 16, 2015**

Vermont Knolls
7900 block of South Normandie Avenue
Age: 51
Gender: Male
Cause: Gunshot
Race/Ethnicity: Black
Agency: LAPD

Follow the story

Counselor shot in Vermont Knolls 'should have never been a target'

POSTED SEPT. 25, 2015, 11:08 A.M.

Todd James Britt, 51

POSTED SEPT. 23, 2015, 3:17 P.M.

Fonte: Reprodução de internet¹¹⁸

¹¹⁷ Texto original: "It is a simple sort of innovation, a simple sort of tool that has strong impact. It does a lot with very little. It is not a complicated piece of code, it is not a complicated theory, it is not a complicated algorithm. It is this simple thing that we built that has impact disproportionate to the amount of time that we actually spend on it. (YOUNG, HERMIDA, 2014)

¹¹⁸ Disponível em: <http://homicide.latimes.com/post/todd-james-britt/> Acessado em 08/09/2017

Esta abertura da plataforma do *Homicide Report* para a interação com a comunidade onde ocorreu o assassinato é considerada pelos jornalistas como um dos aspectos fundamentais deste projeto, como destaca Nicole Santa Cruz, uma das responsáveis pela atualização das notícias.

Homicídios não são iguais, homicídios não são os mesmos. A violência não é a mesma entre as comunidades e entre as raças e, ao expor isso e deixar as pessoas falarem honestamente, a esperança é que talvez elas comecem a pensar de maneira diferente e que você possa estimular uma mudança. (...) Ao expor os níveis de violência, você faz com que as pessoas falem sobre elas de uma maneira real, realista e sem fingimentos.¹¹⁹ (REID, 2014)

Como é possível observar na Figura 19, que ilustra um dos *posts* atualizado por Cruz no *Homicide Report*, além das informações iniciais fornecidas pelo algoritmo, vemos agora a foto da vítima, detalhes do caso, contexto da morte, histórico da vítima e de informações sobre o suspeito, e o telefone da delegacia, caso algum leitor saiba alguma informação e queira contribuir com a investigação. Outro detalhe é a seção “*Follow the story*” na barra lateral esquerda, que exibe o link de uma notícia publicada no *Los Angeles Times* sobre o assassinato em questão.

Ao analisar a percepção dos jornalistas, os pesquisadores Young e Hermida (2014) buscam compreender como o Jornalismo Automatizado foi colocado em prática no *Los Angeles Times*, e como este projeto transformou o modo de pensar desta equipe, modificando normas organizacionais, práticas, estruturas e tecnologias. Uma das conclusões deste estudo observa que, nesta redação jornalística, não houve uma percepção negativa quanto à produção automatizada de conteúdos e eles encaram esta inovação tecnológica como uma ferramenta que os auxilia a atingir a meta de cobrir todos os homicídios da cidade, liberando-os para se dedicarem à apuração e à redação de notícias com maior profundidade.

Os jornalistas do *L.A. Times* entenderam o algoritmo como um reforço para o papel já desempenhado pelos repórteres de polícia, e não como algo que irá substituí-los. Por exemplo, o *L.A. Times* também usa algoritmos para alertar a redação para níveis de crime que excedem o normal para uma área em outro projeto de cobertura de crimes, que não é o *Homicide Report*.¹²⁰ (Idem, p. 9)

Ben Welsh, um dos jornalistas envolvidos no desenvolvimento do *Homicide Report*, destaca em entrevista aos pesquisadores o custo-benefício de se utilizar algoritmos para publicar reportagens sobre pequenos fatos que, por terem um “valor-notícia” baixo, não são

¹¹⁹ Texto original: “Homicide is not equal, homicide is not even. Violence is not even among communities and among races and by exposing that and letting people speak about it honestly the hope is that maybe people start thinking differently and that you could affect change. (...) You expose levels of violence and you get people to talk about them in a way that’s real, realistic and not just pretending.” (REID, 2014)

¹²⁰ Texto original: “Journalists at the L.A. Times understood the algorithm as enhancing the role of crime reporters rather than replacing them. For example, the L.A. Times also uses algorithms to alert the newsroom to levels of crime that exceed the norm for an area in another neighborhood crime project, not the Homicide Report.” (YOUNG, HERMIDA, 2014, p. 9)

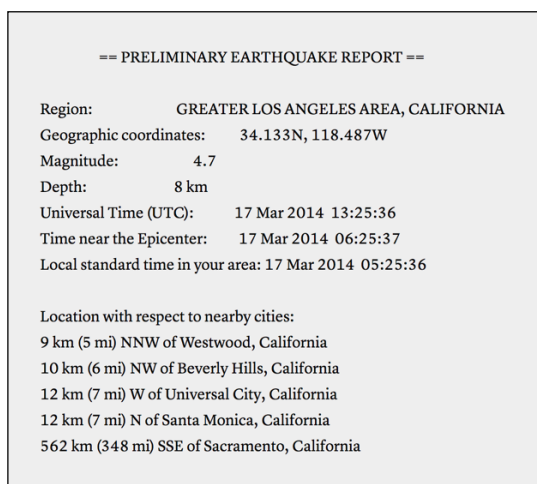
pautados, mas ainda assim são de interesse para uma pequena parcela de leitores. (Ibidem, p. 9) Alguns anos depois, em uma apresentação no NICAR, Welsh refutou a afirmação de Hammond, da *Narrative Science*, sobre a possibilidade de um robô ganhar em breve um prêmio *Pulitzer* e lembra que os computadores têm ajudado os repórteres há anos.

Eu odeio quebrar isso, mas os computadores já ganharam o Prêmio *Pulitzer*. Meia dúzia deles, começando em 1989. (...) O que devemos realmente lutar não é para automatizar por conta apenas da automação ou para economizar dinheiro, porém o que realmente seria ótimo é se pudéssemos facilitar e diminuir a barreira para fazer o tipo de trabalho com o qual se ganha o Prêmio *Pulitzer*.¹²¹ (MARSHALL, 2013)

4.3.2 Quakebot, o robô que escreve (mas não publica) sobre terremotos

No dia 17 de março de 2014, um forte terremoto de 4.7 graus na escala *Ritcher* sacudiu Los Angeles. Apenas oito minutos depois, o jornal *Los Angeles Times* publicou a primeira notícia, através de um *software* que recebeu o nome de *Quakebot* [latimes.com/local/earthquakes] desenvolvido pelo jornalista e programador Ken Schwencke. Identificado pelos equipamentos de monitoramento sísmológico do *United States Geological Survey* (USGS) às 06h25, o terremoto foi reportado a todos os jornalistas às 06h27 em um email com as informações iniciais do evento como mostra a Figura 20. (MEYER, 2014)

Figura 20 – Email enviado pelo USGS com os dados do terremoto



Fonte: MEYER, 2014¹²²

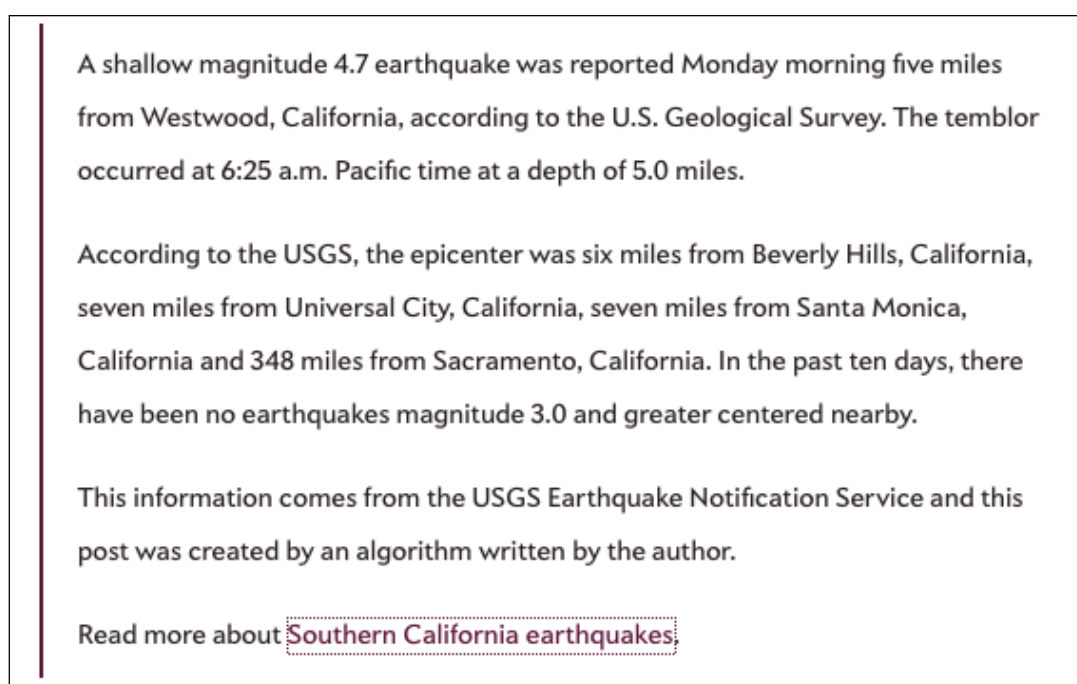
¹²¹ Texto original: "I hate to break it to him but computers have already won the Pulitzer Prize. A half dozen of them, starting in 1989. (...) What we should really strive for is not to automate for automation's sake or to save money, but what would really be great is if we could make it easier and lower the barrier to do the kind of work that wins the Pulitzer Prize." (MARSHALL, 2013)

¹²² Disponível em: <https://www.theatlantic.com/technology/archive/2014/03/how-a-california-earthquake-becomes-the-news-an-extremely-precise-timeline/284506/> Acessado em: 11/10/2017

Foram informações como estas disponibilizadas frequentemente pela USGS para a imprensa que foram utilizadas por Schwencke para desenvolver o *Quakebot*, uma ideia que surgiu em 2011 após um terremoto no Japão. Schwencke pensou em criar um sistema que desmembraria estes dados fornecidos por email para os jornalistas e reescreveria as informações utilizando sentenças simples, o que daria origem a uma notícia automatizada, ou um *robô-post* como ele prefere denominar.

E foi este o sistema utilizado pela primeira vez naquela manhã de março de 2014, o que rendeu o furo ao *Los Angeles Times*, o primeiro jornal a reportar ao público os dados oficiais daquele terremoto e que, por conta da inovação, acabou atraindo a atenção do mundo inteiro.

Figura 21 – Notícia sobre o terremoto escrita pelo *Quakebot*



Fonte: OREMUS, 2014¹²³

Às 6h28, apenas um minuto após receber os dados da USGS, o *Quakebot* enviou um email para os editores do *L.A Times* com a notícia automatizada (Figura 21) e a seguinte mensagem: “Este é um robô-post do seu amigável robô de terremotos. Por favor copie & publique esta notícia.”¹²⁴ O email ainda chama a atenção ao recomendar aos editores que conferissem a cidade destacada no título, em tom irônico e bem humorado.

¹²³ Disponível em: http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html Acessado em: 11/10/2017

¹²⁴ Texto original: “This is a robopost from your friendly earthquake robot. Please copyedit & publish the story.” (MEYER, 2014)

Se a cidade mencionada na manchete é relativamente desconhecida, mas o terremoto aconteceu perto de outra cidade maior, por favor altere o título e o corpo do texto para colocar esta informação primeiro. Eu ainda não sou suficientemente inteligente para tomar estas decisões sozinho e confio na ajuda de seres humanos inteligentes, como vocês.¹²⁵ (MEYER, 2014)

A notícia estava pronta, mas ainda não havia sido publicada. Ao sentir os tremores, Schwencke se levantou e ligou o computador. Ao conferir o texto escrito pelo algoritmo, ele publicou a notícia acompanhada de um mapa com a localização do epicentro às 6h33, oito minutos após os tremores e seis minutos após o algoritmo ter recebido as informações do USGS. Posteriormente, este texto inicial foi atualizado diversas vezes pelo próprio jornalista, inclusive corrigindo a intensidade do terremoto de 4.7 para 4.4 graus, como mostra a Figura 22, em uma reportagem assinada por Schwencke.

Em uma entrevista na época para a *The Wire*, Schwencke exalta a rapidez de se publicar uma notícia em eventos como este com o auxílio de um *software*: *“Quando você tem uma situação como esta e você pode em cinco minutos ter alguma coisa publicada para as pessoas darem um Google e encontrarem... Eu não verifiquei, mas imagino que esta notícia foi muito, muito, muito popular no site hoje.”*¹²⁶ (LEVENSON, 2014)

Schwencke refuta qualquer afirmação que busque associar esta tecnologia à ideia de ameaça ou de um risco futuro de jornalistas serem substituídos por repórteres-robôs. *“A forma como nós a usamos é suplementar. Ela economiza muito tempo e, para certos tipos de notícias, ela transmite a informação de um jeito tão bom quanto se fosse realizado por uma pessoa. Do jeito que eu vejo, ela não elimina o emprego de ninguém da mesma forma que torna o trabalho de todos mais interessante.”*¹²⁷ (OREMUS, 2014)

Em uma entrevista publicada no *YouTube*, Schwencke sustenta que o uso de *softwares* para a produção automatizada de notícias seria uma ferramenta para os jornalistas e argumenta, como um jornalista-programador, que o algoritmo é um reflexo de quem o programa.

Isso afeta a forma como eu programo e o tipo de coisas que eu incluo no código. Como jornalista, eu incluo certas ações e certas condições que refletem a forma como eu analiso o banco de dados. E também a maneira como eu avalio esta informação, principalmente em termos de relevância e

¹²⁵ Texto original: “If the city referenced in the headline is relatively unknown, but the earthquake occurred close to another, larger city, please alter the headline and body text to put that information first. I am currently not smart enough to make these decisions on my own, and rely on the help of intelligent humans such as yourselves.” (MEYER, 2014)

¹²⁶ Texto original: “When you have a situation like that and you can in five minutes have something up there for people to Google and find ... I haven’t checked but I imagine that was a very, very, very popular post on the website today.” (LEVENSON, 2014)

¹²⁷ Texto original: “The way we use it, it’s supplemental. It saves people a lot of time, and for certain types of stories, it gets the information out there in usually about as good a way as anybody else would. The way I see it is, it doesn’t eliminate anybody’s job as much as it makes everybody’s job more interesting.” (OREMUS, 2014)

importância. Como jornalista, eu posso programar isso em um código que, então, será executado por um computador. É uma forma de implementar as minhas ideias. A máquina, de certa forma, está agindo como se fosse eu muito mais do que eu estou agindo como uma máquina.¹²⁸ (FISCHER, 2014)

Figura 22 - Notícia sobre o terremoto após atualizações feitas por jornalista

Los Angeles Times

EDITION: CALIFORNIA | U.S. & WORLD | ENTERTAINMENT | SPORTS | POLITICS | OPINION | MOST POPULAR | TUESDAY, OCT. 10, 2017

Earthquake: 4.4 quake strikes Los Angeles; no damage, injuries reported

This map shows the approximate epicenter of Monday morning's quake near Westwood. (Sino Maps)

By **Ken Schweske**

MARCH 17, 2014, 6:40 AM

A shallow, magnitude-4.4 earthquake was reported Monday morning five miles from Westwood, according to the U.S. Geological Survey. The tremor occurred at 6:25 a.m. PDT at a depth of 5.0 miles.

There were no reports of injuries or damage. The quake was felt over a large swath of Southern California but especially in the Westside and Valley, the *Los Angeles Times* reports.

Updated 6:38 a.m.

The quake was felt over a large area of Southern California. Metro said there might be some delays due to the tremor.

Updated at 6:41 a.m.

The USGS downgraded the quake to a 4.4 magnitude. A spokesman for the *Los Angeles Fire Department* said there were no immediate reports of any damage or injuries. A spokeswoman from the *Los Angeles Police Department* also had no reports of damage.

Updated at 6:46 a.m.

The quake was felt most strongly in the Valley and Westside. According to the USGS "Did You Feel It" service, people reported feeling shaking as far east as Perris and as far south as San Clemente.

Updated at 6:52 a.m.

The LAFD said it is in "earthquake emergency mode as crews survey our City by ground and air; There are NO reports yet of any quake related injuries or damage."

Updated at 6:58 a.m.

California Highway Patrol Officer Monica Posada said there were no reports of any immediate problems on the freeways because of the earthquake. The *Los Angeles County Sheriff's Department* and *County Fire Department* also said there were no reports of immediate damage.

"We did our initial survey and it was felt only. No reports of any damage," said *County Fire Supervisor Michael Pittman*.

Stacey Dirks, the 25-year-old assistant manager at Noah's New York Bagels in Westwood, was at work at the time and said "It just felt like a sudden shake, it was just like rapid and quick."

No bagels fell off the shelves and "everything stayed in place," Dirks said.

Joe Ramallo of the *Los Angeles Department of Water and Power* said that as of 6:45 a.m. there were no reports of water main breaks or power outages immediately following the earthquake. But he said crews are conducting routine safety checks after the quake to ensure their durability.

Updated at 7:06 a.m.

The quake was centered two miles north of Sherman Oaks, not far from the intersection of Mulholland Drive and the 405 Freeway.

According to the USGS, the epicenter was six miles from Beverly Hills, seven miles from Universal City and seven miles from Santa Monica.

In the last 10 days, there have been no earthquakes of magnitude 3.0 or greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

Read more about [Southern California earthquakes](#).

CASA

ADVERTISEMENT

In Case You Missed It

Amid harrowing tales of escape, death toll climbs to 15 as Northern California fires continue to rage
2 hours ago @ 02:55 pm

New disclosure shows a casino guard alerted hotel to gunman before Vegas massacre began. So why did it take so long to stop him?
3 hours ago @ 01:10 pm

24 structures lost and 7,500 acres burned in Orange County firestorm, but slow progress reported
2 hours ago @ 02:55 pm

See More

CHEGOU A HORA DE INVESTIR NOS EUA?

ADVERTISING

Fonte: (SCHWENCKE, 2014)¹²⁹

¹²⁸ Texto original: "It impacts how I code and impacts the kind of things that I put into code. As a journalist there are certain things that I might do or there are certain conditions that I might put into, like how I analyze the data set. And there are certain ways that I might look at something and analyze a piece of information especially for relevance and importance and knowing that. As a journalist I can codify that into a computer code, which then the computer can run. It sort of takes the ideas that I have and implements them for me. A machine is a sort of acting like me more than I am acting like a machine." (FISCHER, 2014)

¹²⁹ Disponível em: <http://www.latimes.com/tn-gnp-earthquake-47-strikes-in-los-angeles-early-monday-2014-0317-story.html> Acessado em: 11/10/2017

No debate em torno do *Quakebot*, o jornalista Robinson Meyer, em uma reportagem para a *The Atlantic*, chama a atenção para uma rede de profissionais e instituições que atua por trás do *software* que escreve automaticamente sobre terremotos.

[...] o [Quakebot] faz parte de um sistema muito maior. Governo e cientistas acadêmicos, técnicos e programadores permitiram o surgimento do *Quakebot*. Eles desenvolveram uma rede de sismógrafos, incluindo instalação, monitoramento e manutenção destes equipamentos. Eles se certificam de que toda a telemetria funcione em caso de emergência, e configuram o sistema para rapidamente distribuir as notícias – e as medidas – de um terremoto. O robô de Schwencke é apenas um nó, um terminal, de um sistema que atinge longas distâncias e recompensa muitas pessoas. *Quakebot*, em outras palavras, não é apenas uma parábola sobre o ‘futuro das notícias’. É uma história sobre dados.¹³⁰ (MEYER, 2014)

Preocupado se o reporter-robô seria uma ameaça para os jornalistas, o jornalista Will Oremus (2014) retoma a polêmica afirmação de Hammond em 2011 e argumenta que esta notícia automatizada “*não é exatamente digna de um Pulitzer – mas as primeiras notícias sobre um terremoto nunca são.*”¹³¹ (OREMUS, 2014) Para Oremus, ao mesmo tempo que o *Quakebot* chama a atenção do mundo para o jornalismo-robô, ele também ilustra suas limitações ao não conseguir acessar os estragos que o terremoto causou na cidade, ao não entrevistar especialistas e também ao não discernir sobre o “valor-notícia” de vários aspectos de um terremoto. Ele também enfatiza que o *Quakebot* não está isento de erros, que podem ocorrer, como Schwencke explica, quando este recebe um alerta falso do USGS. “*Como muitos de seus correspondentes humanos, o Quakebot não verifica os fatos antes de publicar.*”¹³² (OREMUS, 2014)

E foi isso que aconteceu em 21 de junho 2017, quando a USGS enviou erroneamente um alerta sobre um terremoto de 6.8 graus na escala *Richter* que teria atingido Santa Barbara. Fato que, na verdade, havia ocorrido em 1925. Com o alerta, o *Los Angeles Times* acabou publicando automaticamente um *post* no *Twitter*, e teve que se retratar sobre o ocorrido (Figura 23), o que acabou chamando a atenção de muitos jornalistas e leitores.

¹³⁰ Texto original: “it [Quakebot] stands in for a much larger system. Government and academic scientists, technicians, and programmers allowed Quakebot to be born. They developed the networked seismographs, installed, monitored, and maintained them. They made sure all the telemetry works in case of emergency, and set up systems to quickly distribute the news—and measurements—of an earthquake. Schwencke’s bot is only one node, one terminal tendril, of a system that reaches far and rewards many. Quakebot, in other words, isn’t just a “future of news” parable. It’s a story about data.” (MEYER, 2014)

¹³¹ Texto original: “Not exactly Pulitzer-worthy — but then, the first news reports on an earthquake never are.” (OREMUS, 2014)

¹³² Texto original: “Like many of its human counterparts, Quakebot doesn’t double-check its facts before publishing.” (OREMUS, 2014)

Figura 23 – Tweets do L.A. Times se retratando por um erro do Quakebot



Fonte: Reprodução twitter (@LANow e @latimes)

Oremus (2014) destaca que já alcançamos o ponto onde os leitores do *L.A. Times* já não se surpreendem ao se depararem com a frase *“Este post foi escrito por um algoritmo”*¹³³ e pontua que, na sua opinião, a crise que assola o jornalismo com a digitalização em nada tem a ver com as notícias escritas por robôs.

Apenas não espere que os *Quakebots* do mundo forneçam uma contextualização de nível humano ou um julgamento de notícias em breve (...) É verdade que muitos empregos em jornais estão em risco, mas isso não tem nada a ver com os robôs que escrevem notícias. Culpem isso aos modelos de negócios furados por leitores e anunciantes na transição da mídia impressa para a Internet. Seja como for, ajudantes como o *Quakebot* podem salvar alguns empregos ao liberar jornalistas para focarem no tipo de trabalho que só pode ser feito por um repórter local, em campo, e com um cérebro.¹³⁴ (OREMUS, 2014)

4.4 O que os estudos de caso revelam sobre o Jornalismo Automatizado

A cartografia destes três estudos de caso teve como intuito abrir as caixas-pretas e evidenciar algumas das redes sociotécnicas por trás da produção automatizada de notícias, em busca de um ponto de vista simétrico que não supervalorizasse o lado não-humano desta inovação tecnológica e que conseguisse reconhecer que a dimensão humana ultrapassa a

¹³³ Texto original: “This post was written by an algorithm.” (OREMUS, 2014)

¹³⁴ Texto original: “Just don’t expect the Quakebots of the world to provide human-level context or news judgment anytime soon (...) It’s true that a lot of newspaper jobs are in danger, but that has nothing to do with news-writing robots. Blame that on a business model punctured by readers’ and advertisers’ move from print to the Internet. If anything, helpers like Quakebot might save a few the ground, with a brain.” (OREMUS, 2014)

etapa de programação dos algoritmos. Ao longo da pesquisa, observamos que muitas vezes parte destas redes sociotécnicas são invisibilizadas por um discurso que associa esta tecnologia à imagem de robôs humanoides e à ideia de que, no futuro, notícias escritas por computadores ganharão o prêmio *Pulitzer*.

Um dos primeiros pontos desta cartografia que nos chama a atenção é o **discurso mercadológico** assumido pelos empreendedores tanto da *Narrative Science* quanto da *Automated Insights* no momento inicial do lançamento das *startups*. Neste período, o interesse destes executivos era atrair visibilidade para conquistar novos clientes e monetizar o projeto que, nos idos dos anos 2010 e 2011, ainda estava mais no campo das ideias do que de uma aplicação prática efetiva. Em ambas *startups*, o foco era na implementação dos *softwares* NLG para a cobertura esportiva e, entre algumas declarações, vemos por exemplo Kristian Hammond, um dos fundadores da *Narrative Science*, afirmando que em quinze anos mais de 90% das notícias seriam escritas por um computador, e não por um repórter. Porém, ele acreditava que isto não deveria ser motivo de alarde para nenhum jornalista, pois o “tsunami de robônotícias” não iria eliminar o emprego de ninguém e sim “expandir” a cobertura. (LEVY, 2012) Em um discurso semelhante, Robbie Allen na mesma época afirmava que um de seus objetivos com a *StatSheet* era que 80% dos leitores não percebessem que o conteúdo não havia sido escrito por um jornalista, e sim por um *software*. (STROSS, 2010) Apesar de exaltar a capacidade técnica de seus algoritmos, Allen corrobora com a opinião de Hammond de que esta inovação não iria retirar o emprego de ninguém (GRIMM, 2010), e reproduz um discurso recorrente de que esta tecnologia liberaria os jornalistas para executarem tarefas mais interessantes e menos repetitivas.

Se por um lado temos neste período inicial um grupo de empreendedores disseminando uma visão determinista e megalomaniaca que acredita na produção maquinica de um grande volume de notícias, por outro, quando olhamos para os textos escritos pelos *softwares* NLG, nos deparamos com narrativas simples e encadeadas por dados estatísticos, que reproduzem uma estrutura repetitiva e formulaica. Neste contexto, um questionamento que nos parece central nesta discussão emerge: qual é a visão de jornalismo destes empreendedores? Será que Hammond, em 2011, acreditava que estas notícias hiperlocais com a cobertura de jogos universitários poderiam um dia ganhar o *Pulitzer*?

Ao analisarmos esta cartografia, nos parece que as declarações de Hammond, Allen e Stuart Frankel estão alinhadas a um discurso mercadológico sustentado por uma visão de jornalismo simplificada e superficial, que se limita à produção de pequenas notas dirigidas por dados. Neste discurso, não há o reconhecimento do papel social e político desempenhado pelo jornalismo, e nem do trabalho minucioso, aprofundado e detalhista realizado por jornalistas que já foram condecorados com o *Pulitzer*.

Neste cenário, quando a *Narrative Science* levanta a hipótese de que um dia sua tecnologia seria agraciada com um *Pulitzer*, nos parece que esta foi uma estratégia para atrair visibilidade e conferir legitimidade ao trabalho por eles executado, como um selo de qualidade que agregaria valor ao produto final, no caso, às notícias produzidas pelo *software Quill*. E nos parece que eles tinham a consciência da controvérsia que isto iria causar, mas queriam incitar um debate para atrair a atenção para a tecnologia que estavam desenvolvendo. E em alguma medida isto deu certo. Um dos fatos que sustenta nossa avaliação de que isto fazia parte da estratégia foi uma recente declaração de Stuart Frankel, CEO da *Narrative Science*, em março de 2018 para o *MIT Technology Review*: *"Eu sempre brinco que se você quiser atrair a imprensa em um estágio inicial da empresa, faça algo que se pareça disruptivo para o jornalismo, pois repórteres amam escrever sobre a sua própria indústria."*¹³⁵ (WINICK, 2018)

É interessante pontuar ainda que Robbie Allen, CEO da *Automated Insights*, cita o *Pulitzer* mesmo antes da polêmica declaração de Hammond ao *The New York Times*, quando afirma em 2010 que *"o conteúdo [produzido por seu software] pode ser um pouco duro em algumas partes, mas você não tem muitos vencedores do Pulitzer escrevendo sobre esportes, então o nível não é muito alto."* (GRIMM, 2010) As inúmeras vezes que o *Pulitzer* é citado no debate em torno do Jornalismo Automatizado só demonstram o quão importante este prêmio é para o reconhecimento de um trabalho jornalístico de qualidade, pelo menos nos Estados Unidos.

Um outro ponto que nos parece suscitar uma discussão interessante neste período inicial do Jornalismo Automatizado é o **conceito de "audiência para um"**, defendido tanto por Frankel quanto por Allen. Em 2012, o CEO da *Narrative Science* declara que uma das suas principais motivações era criar notícias sobre assuntos raramente cobertos por jornalistas e defende o uso desta tecnologia para a produção de notícias que seriam lidas por uma ou poucas pessoas. (FASSLER, 2012) Com um discurso semelhante, em 2014 o CEO da *Automated Insights* afirmou que o Jornalismo Automatizado seguia uma outra lógica onde, ao invés de se produzir notícias que seriam lidas por milhares de pessoas, o objetivo seria criar um milhão de artigos para serem lidos, cada um, por apenas uma pessoa. (KIRKLAND, 2014)

Esta associação dos *softwares* NLG a um esforço de se personalizar a cobertura jornalística parece ir de encontro a uma demanda pela diferenciação dos conteúdos produzidos pelos jornais nas plataformas digitais, contrariando a "espiral da mesmice" (BOCZKOWSKI, 2010) que identifica uma tendência à imitação e homogeneização nestas

¹³⁵ Texto original: "I always joke that if you want to get a lot of press as an early-stage company, do something that is perceived to disrupt journalism, because reporters love to write about their own industry." (WINICK, 2018)

coberturas. Por outro lado, se considerarmos que estas notícias “personalizadas” são produzidas seguindo um mesmo *template*, esta demanda pela diferenciação pode não ocorrer de fato já que milhares de notícias serão publicadas seguindo uma mesma estrutura. Assim, estamos diante de um paradoxo pois, se num primeiro olhar, pode parecer que este conceito de “audiência para um” vai de encontro a uma crescente diferenciação e se opõe à ideia de “espiral da mesmice”, ao reconhecermos as semelhanças dos textos produzidos podemos afirmar que, ao final, o Jornalismo Automatizado eleva exponencialmente o grau de imitação e homogeneização do jornalismo digital.

Quando a *Narrative Science* anunciou sua parceria com a *Forbes*, em uma das entrevistas Frankel afirma que a aplicação de *softwares* NLG nas redações criaria um **novo profissional**, o “meta-jornalista”, responsável pela redação de milhões de notícias ao mesmo tempo, e não apenas uma. (FASSLER, 2012) Na prática, nos parece que esta ideia não foi confirmada pois, quando avaliamos a experiência da *startup* com a *ProPublica*, vemos que os desafios de produzir 52 mil notícias de uma única vez fugiram um pouco do esperado. Sobre este projeto, Hammond reconhece, por exemplo, as dificuldades que enfrentaram onde, apesar dos dados parecerem semelhantes, as várias condições ambientais acabaram por criar falsas analogias e muitas das comparações, simplesmente, não funcionaram. (LECOMPTE, 2015) O jornalista da *ProPublica* Scott Klein também destacou, na época, o desafio de editar 52 mil notícias ao mesmo tempo: “*Editar uma notícia não significa que você editou todas elas.*” (KLEIN, 2013) Ele explica que, em muitos casos, uma edição que fazia sentido em um contexto não funcionava em outros e, para ele, a produção automatizada de notícias pareceu ser uma “*solução intrigante*” para o problema de escalonamento enfrentado pelo jornalismo na era do *big data*.

O projeto *The Opportunity Gap* da *ProPublica*, realizado em parceria com a *Narrative Science*, é um exemplo interessante de aplicação de *softwares* NLG para a cobertura de uma temática de maior interesse público, no caso a educação. Neste exemplo, assim como no *The Homicide Report* do *Los Angeles Times*, vemos uma aproximação do Jornalismo Automatizado com o Jornalismo de Dados, onde os algoritmos auxiliam no processamento de um grande volume de informações que criam formatos de cobertura diferenciados. Nestes dois casos temos a publicação do conteúdo em uma estrutura mais elaborada, onde a notícia automatizada é um dos elementos de um site com outras funcionalidades, como filtros, ferramentas de busca e visualizações de gráficos, mapas, dados geolocalizados.

Ao observarmos a trajetória da *Narrative Science* e da *Automated Insights*, é possível traçar um paralelo e observar muitas semelhanças entre estas duas *startups*. Ambas começaram com um protótipo para a cobertura automatizada de jogos universitários e se consolidaram no mercado através de uma parceria com um veículo jornalístico de maior

alcance, a *Forbes* e a *Associated Press* respectivamente, para a produção de balancetes financeiros de empresas listadas na bolsa. Estas duas empresas travaram uma disputa competitiva no mercado durante alguns anos, até a *Narrative Science* desistir do jornalismo. Um dos fatores que corroboram com esta análise é, por exemplo, o fato dos balancetes financeiros tanto da *Forbes* quanto da *Associated Press* usarem como fonte de dados a mesma empresa, a *Zacks Investment Research*.

Na cartografia, uma outra questão que aparece é a **da autoria de conteúdos produzidos por software**, uma importante discussão ética em torno do Jornalismo Automatizado. Enquanto os conteúdos tanto da *Forbes* quanto da *ProPublica* indicam a fonte de dados e o nome da *Narrative Science*, o mesmo não acontece nos sites da *StatSheet*, e Allen confirma que nem todos os clientes da *Automated Insights* revela que suas notícias foram escritas por um “robô” ou *software* de Inteligência Artificial. Por mais que estes clientes de Allen tenham que seus conteúdos tenham um tratamento diferenciado por sistemas de busca, defendemos que um dos pontos fundamentais nesta discussão é a transparência, tanto para pontuar se um conteúdo foi escrito por um *software* e qual a fonte de dados utilizada, quanto para informar como estes *softwares* funcionam e evitar o encaixapretamento de processos vinculados ao Jornalismo Automatizado. No caso da *Associated Press* e do *Quakebot*, a autoria algorítmica é indicada pelo texto, mas o mesmo não acontece no *The Homicide Report*, por exemplo.

Mudando o foco das empresas e dos empreendedores para os jornalistas, uma das opiniões que se destaca na cartografia é a do repórter Joe Fassler, da revista *The Atlantic*. Decidido a investigar o assunto, ele revela que chegou à sede da *Narrative Science* com a expectativa de ter suas próprias convicções confirmadas: a de que a mente humana é um “mistério sagrado” e que a relação que os escritores e jornalistas têm com as palavras é “única e profunda”, duvidando da capacidade de escrita de um algoritmo. “*Mas conversando com Hammond, eu percebi o quanto o processo da escrita – o que eu acreditava que era imprevisível, até variável – pode ser quantificado e modelado*”, afirmou Fassler (2012). De certa forma, parece que Fassler ficou surpreso com a performance dos *softwares* NLG, assim como David Carr (2009) do *The New York Times* que estranhou o “*fato das notícias não serem terríveis*” e Cathal Kelly (2009) do *Toronto Star* que se surpreendeu por não encontrar “*nenhum erro aparente de apuração ou de gramática*”.

Olhando especificamente para a experiência da *Associated Press* com o Jornalismo Automatizado, conseguimos compreender melhor como se dá a **atuação dos jornalistas e editores** no processo de produção de uma notícia utilizando *softwares* NLG. Vemos, por exemplo, a partir do discurso da editora-assistente Philana Patterson, que jornalistas trabalharam junto com programadores para transformarem as regras e os modelos de notícias

sobre balancetes financeiros em código algorítmico. (LECOMPTE, 2015) Há ainda uma afirmação recorrente, tanto no discurso de Patterson quanto no da editora Lisa Gibbs e do vice-presidente Lou Ferrara, de que esta tecnologia liberaria os jornalistas de tarefas rotineiras e entediadas para focarem em pesquisas mais aprofundadas e na produção de “notícias mais complexas”. (MARCONI, SIEGMAN, 2017) Nesta perspectiva, eles tentam associar o uso destes *softwares* a ferramentas que auxiliam a cobertura das temporadas de resultados, o que ampliou a produção de notícias de 440 para quase cinco mil artigos sem que isso sobrecarregasse a equipe de repórteres.

Um dos aspectos da experiência da *Associated Press* que se destaca é a criação do cargo de “editor de notícias automatizadas”, que teria entre suas atribuições identificar tarefas repetitivas e onerosas desempenhadas pelos jornalistas com potencial para serem automatizadas. O profissional responsável pela função, Justin Myers, tem uma visão particular sobre a qualidade dos conteúdos automatizados, que para ele seriam o resultado de “explicação e teste”. (MARCONI, SIEGMAN, 2017) A “explicação” estaria relacionada à capacidade dos editores detalharem os processos que transformam dados estruturados em um texto jornalístico. E o “teste” estaria relacionado à avaliação dos conteúdos produzidos, tanto em uma fase *beta* de desenvolvimento do *software* quanto em um posterior monitoramento, quando este sistema já está em funcionamento. Para Myers, “explicação e teste” seriam fundamentais tanto para fins de “transparência editorial” quanto para a resolução de problemas, quando novamente a equipe multidisciplinar de editores, programadores e “técnicos de implementação” seriam acionados para aprimorarem o sistema.

Se contrapondo à visão de Robbie Allen, CEO da *Automated Insights*, de que no Jornalismo Automatizado “o único envolvimento humano era com a criação do algoritmo que gera as postagens” (SCHONFELD, 2010), nos parece que a *Associated Press* defende que a aplicação de Inteligência Artificial no jornalismo abre oportunidades para profissionais de outras áreas, como programadores e cientistas de dados. (MARCONI, SIEGMAN, 2017) Eles possuem em sua equipe de jornalismo de dados, por exemplo, o matemático Larry Fenn que defende que as redações valorizem mais as ciências exatas, o que ele chama de “boa ciência”, repercutindo uma visão tradicional de conhecimento científico que é bem criticada tanto pelo STS quanto pela TAR.

Em um contraponto com os dois primeiros estudos de caso que estão centrados em parcerias entre *startups* de tecnologia e empresas de mídia, quando nos voltamos para o exemplo do *Los Angeles Times*, é interessante ressaltar que a adoção do Jornalismo Automatizado partiu de uma demanda interna da redação e como uma iniciativa dos próprios jornalistas que também atuaram como programadores. Em 2010, na mesma época que a *Narrative Science* e a *Statsheet* focavam na cobertura de jogos universitários, o *L.A. Times*

adaptava esta tecnologia para viabilizar um projeto já existente no jornal de reportar todos os casos de homicídios da região metropolitana de Los Angeles, nos Estados Unidos. Semelhante ao *The Opportunity Gap*, o *The Homicide Report* focava em um assunto de maior interesse público e tinha como missão ampliar a cobertura e dar visibilidade para os casos mais frequentes envolvendo jovens negros que normalmente não recebiam a atenção da mídia. Num primeiro momento, em 2007, o jornal tentou desenvolver o projeto com uma repórter, mas o volume de casos acabou inviabilizando o projeto. Foi neste contexto que a equipe de jornalistas de dados do *L.A. Times* desenvolveu o sistema que publicaria automaticamente um “robôpost” para cada homicídio reportado pela *Los Angeles County coroner’s office*, sendo que parte deste conteúdo seria atualizado por repórteres, adicionando mais detalhes de alguns casos.

Esta lógica de notícias automatizadas sendo atualizadas por jornalistas após a sua publicação também foi seguida pelo outro projeto do *L.A. Times*, o *Quakebot*, o que nos leva a crer na existência de uma relação mais integrada entre os jornalistas e a cobertura feita pelos *softwares* NLG. Como aponta Young e Hermida (2014), os jornalistas do *L.A. Times* encararam esta tecnologia como uma ferramenta que os auxiliaria a ampliar a cobertura de homicídios e reportar de forma mais rápida os terremotos, e não como algo que poderia substituí-los. É interessante pontuar que o editor da equipe de jornalismo de dados, Ben Welsh, foi um dos que criticou a declaração de Hammond sobre o prêmio *Pulitzer*. Contra o uso da automação para “economizar dinheiro”, ele defende que os computadores já ganharam vários *Pulitzer* desde 1989, e que esta tecnologia deve ser adotada não para substituir repórteres, e sim como um suporte para se produzir um jornalismo de qualidade. (MARSHALL, 2013)

Diferente do discurso megalomaniaco dos executivos da *Narrative Science* e da *Automated Insights*, que sempre supervalorizaram a capacidade dos algoritmos, o jornalista-programador Ken Schwencke, envolvido no desenvolvimento dos dois projetos do *L.A. Times*, assume uma postura muito mais humilde ao afirmar que os *softwares* NLG são “um tipo simples de inovação, um tipo simples de ferramenta que tem um forte impacto”. (YOUNG, HERMIDA, 2014) Ele pontua que não há nada de complicado no algoritmo desenvolvido, e em um outro momento ainda esclarece que o *software* reproduz os valores, as ideias e a visão de quem o programa. “Como jornalista, eu incluo certas ações e certas condições que refletem a forma como eu analiso o banco de dados. [...] A máquina, de certa forma, está agindo como se fosse eu muito mais do que eu estou agindo como uma máquina.” (FISCHER, 2014)

Por fim, um último ponto de análise desta cartografia é sobre **situações onde o software NLG erra**, o que ocorre na maior parte das vezes por problemas nos bancos de dados vinculados a estes sistemas. No caso do *Quakebot*, o erro ocorreu por conta de um alerta falso do USGS que recuperou os dados de um terremoto ocorrido em 1925. Diferente

da ocorrência de março de 2014, em 2017 o *software* não mandou apenas um alerta para o email dos editores do *L.A. Times*, mas publicou instantaneamente um *tweet* assim que recebeu as informações. Como o alerta era falso, o jornal teve que se retratar publicamente do ocorrido e acabou atraindo a atenção para o erro dos “robôs”. Este caso corrobora com a opinião de Philana Patterson e Lisa Gibbs da *Associated Press*, que frisaram em suas declarações que os erros ocorrem quando o *input* de dados está incorreto. “Se os dados estiverem ruins, você terá uma notícia ruim.”¹³⁶ (MARCONI, SIEGMAN, 2017)

De modo geral, consideramos que a descrição detalhada destes três estudos de caso nos auxiliou a compreender de maneira mais aprofundada o contexto no qual o Jornalismo Automatizado está sendo implementado nas redações, sanando algumas dúvidas e incertezas que normalmente são associadas a esta temática, e levantando outras questões acerca do uso crescente de algoritmos nas práticas jornalísticas. Voltando o nosso olhar para as redes sociotécnicas, nos parece que este mapeamento elucida em muitos pontos a forma como *softwares* NLG são adotados para a produção de notícias automatizadas. Considerar estas narrativas como resultado apenas de uma ação maquínica, como muitos defendem e acreditam, é uma forma simplificada de se olhar para esta inovação que encaixapreta um complexo agenciamento de atores humanos e não-humanos que possibilitam o funcionamento de sistemas como esse.

Assim, os estudos de caso desta cartografia nos fornecem elementos que nos permitem afirmar que a **dimensão humana por trás do Jornalismo Automatizado** extrapola a atuação apenas de programadores limitada a uma etapa inicial de programação de algoritmos. Vimos nestas descrições vários jornalistas e editores envolvidos nestes projetos, desde a etapa de desenvolvimento como destacado por Philana Patterson da *Associated Press*, passando pelo desafio de se editar milhares de notícias como pontuado por Scott Klein da *ProPublica*, e chegando ao processo de atualização dos conteúdos automatizados por repórteres, como rotineiramente acontece no *Los Angeles Times*. Percebemos ainda neste mapeamento que as notícias automatizadas estão bem próximas do Jornalismo de Dados, vinculadas ao trabalho de equipes multidisciplinares que envolvem, além de jornalistas e programadores, profissionais de outras áreas como cientistas de dados, matemáticos e empreendedores.

¹³⁶ Texto original: “If the data’s bad, you get a bad story.” (MARCONI, SIEGMAN, 2017)

5 Considerações finais

Ao olhar para o Jornalismo Automatizado, um objeto que nos desafia a refletir sobre a crescente mediação algorítmica da profissão, esta dissertação se propôs a traçar as redes sociotécnicas vinculadas a esta inovação tecnológica e compreender as controvérsias e incertezas desencadeadas por uma “suposta” autonomia tecnológica de “robôs” redatores de notícias. Com uma fundamentação teórica ancorada no STS e na Teoria Ator-Rede, cartografamos pesquisas, experimentos e três estudos de caso com o intuito de revelar parte dos agenciamentos entre atores humanos e não-humanos que atuam na produção de notícias automatizadas.

Buscamos nesta pesquisa um ponto de vista simétrico que não exaltasse nem uma perspectiva determinista e utópica que se dirige aos avanços tecnológicos a partir de um discurso revolucionário, nem uma visão pessimista que associasse esta inovação à ideia de ameaça e medo, em um futuro onde as redações seriam tomadas por robôs humanoides. Nos inspiramos no STS na busca por um discurso de “desconstrução” (AHVA, STEENSEN, 2017) que estivesse atento para as sócio-materialidades relacionadas a esta tecnologia, para a formação de redes e para um entendimento sobre como o *big data*, os algoritmos e a Inteligência Artificial vem sendo incorporados ao jornalismo digital.

É importante pontuar que o jornalismo sempre teve suas atividades e práticas intrinsicamente vinculadas a diversas tecnologias, e que a automação está presente na profissão há pelo menos quatro décadas (LINDEN, 2016). Porém, como destaca Boczkowski (2004), existe um conservadorismo no jornalismo que colabora para a disseminação de uma visão pessimista frente à introdução de novas tecnologias, onde os jornalistas são historicamente condicionados a reagir em prol da manutenção de processos tradicionais e já cristalizados, e demonstram dificuldades para se adaptarem às inovações.

Por outro lado, apesar da automação não ser uma novidade, é necessário reconhecer que desde 2010 o Jornalismo Automatizado tem sido alvo de controversos debates entre jornalistas e leitores por conta de projetos como os desenvolvidos pelas *startups Narrative Science* e *Automated Insights*, por veículos como a *Forbes*, a *ProPublica*, a *Associated Press* e o *Los Angeles Times*. Sobre a aplicação de *softwares* NLG em atividades jornalísticas, a cartografia realizada por esta pesquisa sinaliza para **um uso restrito desta inovação** em redações online para a produção de notas e notícias simples e repetitivas de temáticas específicas como finanças, esportes, eleições, crimes, terremotos, previsões do tempo. Esta tecnologia também foi adotada em projetos vinculados ao Jornalismo de Dados com temáticas de maior interesse público, como o *The Opportunity Gap* da *ProPublica* e o *The Homicide Report* do *Los Angeles Times*. Estes exemplos demonstram um potencial para esta tecnologia

ser utilizada pelo jornalismo em pautas onde seja necessário o processamento de um grande volume de dados.

É recorrente ainda a associação do Jornalismo Automatizado ao fenômeno da “cauda longa” como uma estratégia para ampliar a cobertura e aumentar a audiência com o uso de algoritmos. Uma das promessas evocadas por executivos e empresários é a de que estes sistemas possibilitariam uma crescente personalização de conteúdos, que seriam dirigidos a um público de nicho, e não seguiriam uma lógica de comunicação de massa. Este discurso levanta uma preocupação sobre o quanto o Jornalismo Automatizado contribuiria para o processo de **comoditização das notícias** (CARLSON, 2014).

Como nos mostra a cartografia realizada nesta pesquisa, as notícias automatizadas são um fenômeno situado em um contexto de digitalização das práticas jornalísticas e, da mesma forma que Boczkowski (2010) sugere que há um crescente grau de imitação e homogeneização das notícias no jornalismo digital, o que ele chama de “espiral da mesmice”, esse fenômeno adquire proporções ainda maiores quando *softwares* NLG são inseridos nas redações. Estes sistemas são configurados para produzirem e publicarem um grande volume de narrativas a partir de templates formulaicos e dados estatísticos. Na prática, o que ocorre é a publicação de milhares de notícias com estruturas simples e repetitivas, semelhantes entre si e que se diferenciam no sujeito em destaque e no processamento das variáveis numéricas.

Nesta pesquisa, sugerimos que o Jornalismo Automatizado seja um campo de estudos em formação que dialoga com as Ciências da Computação e com o Jornalismo Computacional, e que se estabelece a partir de uma **forte articulação com universidades, startups, patentes e empresas de mídia**.

Na academia, uma das principais preocupações dos pesquisadores é quanto à recepção destas notícias automatizadas. Estariam os leitores aptos a distinguir se um texto foi escrito por um algoritmo ou por um repórter? Algumas pesquisas recentes destacam que eles têm dificuldades de pontuar a autoria dos textos, exaltando a qualidade dos conteúdos produzidos por *software* NLG. Estas percepções comumente avaliam as notícias automatizadas como mais objetivas e descritivas, enquanto as escritas por jornalistas como mais coerentes e prazerosas de ler. (CLERWALL, 2014) De certa forma, estas pesquisas nos surpreenderam por utilizarem metodologias clássicas que comparam conteúdos escritos por repórteres e por softwares, dando origem a gráficos analíticos em um formato que não é usual para os estudos de jornalismo. Estas pesquisas demonstram um esforço acadêmico de se **verificar a legitimidade e a eficiência destes “robôs” que redigem notícias**, na busca por reflexões que ajudem a compreender os impactos desta tecnologia no jornalismo.

Em meio a tantas incertezas, um dos fatores que nos motivou a conduzir esta investigação foi a **perspectiva tecnocêntrica de alguns pesquisadores** que enfatizam que

“o elemento humano direto e ativo é eliminado do processo de criação no Jornalismo Algorítmico” (DÖRR, 2015) ou que estas notícias são resultado de “processos algorítmicos [...] com ação limitada ou nenhuma intervenção humana além das escolhas iniciais de programação” (CARLSON, 2014). No nosso ponto de vista, esta parece ser uma visão simplificada do Jornalismo Automatizado que prioriza suas análises nos textos finais publicados pelos veículos, mas que invisibiliza os processos e a complexa rede de profissionais e tecnologias mobilizadas por esta inovação.

Estas definições parecem superestimar o papel desempenhado pelos programadores, os colocando em um lugar central na dimensão humana por trás dos softwares NLG que não se atenta para o trabalho desempenhado por outros profissionais, como jornalistas, editores, empresários e analistas de dados. Além disso, defendemos que a adoção destes sistemas por uma redação passa por pelo menos quatro etapas - desenvolvimento, implementação, monitoramento e manutenção - sendo que todas elas exigem o trabalho de uma equipe transdisciplinar cercada de muitos artefatos técnicos. Este é um dos pontos interessantes levantados por esta pesquisa, que podem ser explorados com mais profundidade em trabalhos futuros.

Para fundamentar este ponto de vista em busca das redes sociotécnicas vinculadas ao Jornalismo Automatizado, duas concepções de autonomia tecnológica foram essenciais. De acordo com a perspectiva de Winner (1977), os “artefatos técnicos do século XX” são caracterizados pelo “enorme tamanho, complexas interconexões e uma interdependência sistêmica”, onde uma parte é essencial para o trabalho de outras partes. Sugerindo uma visão que nos parece complementar, Latour (1994b) estabelece a delegação como um dos significados de mediação técnica, e considera a tecnologia como um “trabalho congelado”, híbridos que trazem para o presente várias ações passadas de outros actantes, humanos e não-humanos que continuam agindo nos objetos que ajudaram a confeccionar. Estes dois conceitos nos ajudam a reconhecer a complexidade das redes sociotécnicas que dão origem aos artefatos técnicos e, mesmo que invisíveis, não podemos desconsiderar a atuação dos profissionais que desenvolveram os *softwares* NLG quando nos deparamos com uma notícia automatizada.

Inspirados por estas perspectivas e pela cartografia tanto das pesquisas quanto dos três estudos de caso, argumentamos que **existe um novo ecossistema em formação** onde cada vez mais jornalistas trabalham em equipes multidisciplinares, ao lado de engenheiros, programadores, analistas de dados, envolvidos ainda com universidades, pesquisadores, startups e empreendedores. Sugerimos ser necessário aprofundar as discussões em torno do Jornalismo Automatizado para que se reconheça a complexidade das redes sociotécnicas vinculadas a esta inovação, que não podem ser simplificadas com a disseminação da ideia

de textos escritos por algoritmos sem nenhuma intervenção humana. Por exemplo, a cartografia desta pesquisa nos sinaliza como jornalistas atuam na etapa de desenvolvimento destes *softwares* ao lado de programadores, como destacado por Philana Patterson da *Associated Press*, a dificuldade de se editar milhares de notícias de uma vez como relatado por Scott Klein da *ProPublica*, como repórteres atualizam as notícias automatizadas como acontece no *Los Angeles Times*.

A automação das práticas jornalísticas é acompanhada por uma crescente complexificação dos processos de produção de notícia, em um cenário onde softwares, algoritmos e big data introduzem novos desafios e oportunidades para a profissão. Porém, a implementação desta tecnologia nas redações exige um olhar atento que não ignore os impactos sociais e políticos e, neste cenário, **novas discussões éticas emergem**. Por exemplo, um dos pontos que nos parece fundamental de ser considerado é a questão da transparência no que tange o uso de algoritmos e *big data* para fins jornalísticos. Tanto repórteres quanto leitores devem estar cientes se uma notícia foi escrita por um jornalista ou por um *software*, e que tipos de dados foram utilizados como input para a produção dos conteúdos automatizados. As empresas precisam, por outro lado, se certificar de que possuem os direitos legais para o uso destes dados, que eles são precisos e provêm de uma fonte confiável. Em um artigo onde discute a autoria e a responsabilidade legal das notícias automatizadas, Montal (2016) sugere que estes conteúdos sejam creditados aos profissionais responsáveis por sua criação, o que envolve tanto jornalistas e editores quanto programadores. Mas e as empresas que adotam estes sistemas? Não seriam elas também responsáveis?

Entre as **limitações desta pesquisa**, temos consciência de que nossa cartografia é limitada às escolhas tomadas ao longo deste processo, e que as descrições revelam apenas algumas redes sociotécnicas vinculadas ao Jornalismo Automatizado, em um relato subjetivo que reflete o modo como olhamos para este objeto. Uma das principais limitações desta pesquisa é o foco para experimentos e estudos de caso relatados em inglês, o que acaba por priorizar exemplos onde esta tecnologia é adotada nos Estados Unidos. Se por um lado, reconhecemos o pioneirismo tanto das experiências quanto das pesquisas que são evidenciadas em nosso mapeamento, por outro sabemos que este campo é muito mais amplo e está em plena expansão. Mas por conta de barreiras linguísticas e geográficas, tivemos ao longo desta pesquisa que fazer escolhas que limitassem a empiria a este universo, apesar do desejo genuíno de continuarmos em busca de mais informações sobre a aplicação de *softwares* NLG no jornalismo, tanto no Brasil quanto no mundo.

Consideramos importante pontuar que as pesquisas futuras devem estar atentas ao modo situado com que tecnologias como os *softwares* NLG se inserem nas redações,

alinhado ao conceito de “mídias emergentes” sugerido por Boczkowski (2004). Esta perspectiva defende que as inovações acontecem de forma gradual e são moldadas por infraestruturas já existentes. Assim, pesquisas futuras devem buscar as especificidades de cada contexto e enfatizar as questões históricas, locais e processuais associadas a inserção de uma nova tecnologia nas práticas jornalísticas.

Quanto ao futuro, defendemos que é necessário que os jornalistas encarem o potencial do *big data* e de sistemas de Inteligência Artificial para o avanço técnico da profissão, auxiliando-os a lidar com o crescimento exponencial de dados e informações disponíveis atualmente no mundo. A implantação de *softwares* NLG nas redações desafia os jornalistas a lidarem com a apuração de uma nova forma, e uma tendência talvez seja a necessidade deles se familiarizarem com linguagens de programação, principalmente *Python* e *R*. A partir dos estudos de caso, visualizamos oportunidades onde o Jornalismo Automatizado pode ser adotado em projetos com temáticas de maior interesse público que necessitem processar um grande volume de dados, em um encontro com o Jornalismo de Dados.

Referências Bibliográficas

- AHVA, Laura; STEENSEN, Steen. Deconstructing digital journalism studies. IN: FRANKLIN, Bob; ELDRIDGE II, Scott A. (Ed.). **The Routledge Companion to Digital Journalism Studies**. Abignon, Oxon: Routledge, 2017. p. 25-34
- AL JAZEERA ENGLISH. **Robot Journalism: the end of human reporters?** Vídeo (9m13s) 15/11/2015 Disponível no Youtube: <https://www.youtube.com/watch?v=ci-rHRJIFew> Acessado em: 12/10/2017
- ALLEN, Nicholas D.; TEMPLON, John R.; MCNALLY, Patrick S.; BIRNBAUM, Larry. HAMMOND, Kristian. StatsMonkey: A Data-Driven Sports Narrative Writer. In: **AAAI Fall Symposium: Computational Models of Narrative**. 2010. p. 2-3.
- ANDERSON, Christopher W. Towards a sociology of computational and algorithmic journalism. **New Media & Society**. n. 15, v. 7, p. 1005-1021. 2012. DOI: 10.1177/1461444812465137
- ANDERSON, Chris. **A cauda longa**. Elsevier Brasil, 2006.
- AP ARCHIVE. **Automated Journalism in the 21st century newsroom**. Vídeo (5m09s) 03/08/2015. Disponível em: https://www.youtube.com/watch?v=TmGmN8_yJAM Acessado em: 12/10/2017
- BARBOSA, Suzana. Modelo JDBD e o ciberjornalismo de quarta geração. IN: FLORES VIVAR, J. M. RAMÍREZ, F. E. (ed.) **Periodismo Web 2.0**. Madrid: Editorial Fragua, 2009. p. 271-283.
- _____. **Jornalismo Digital em Base de Dados (JDBD)** Um paradigma para produtos jornalísticos digitais dinâmicos (Tese de Doutorado). FACOM/UFBA, Salvador, 2007. Disponível em: http://www.facom.ufba.br/jol/producao_teses.htm
- BENSON, Rodney. From Heterogeneity to Differentiation: Searching for a Good Explanation in a New Descriptivist Era. IN: BOCZKOWSKI, Pablo J. ANDERSON, C. W. **Remaking the News: Essays on the Future of Journalism Scholarship in the Digital Age**. MIT Press, 2017. p. 27-45.
- BERCOVICI, Jeff. **Can you tell a robot wrote this? Does it matter?** Forbes 29/11/2010. Disponível em: <https://www.forbes.com/sites/jeffbercovici/2010/11/29/can-you-tell-a-robot-wrote-this-does-it-matter/> Acessado em: 07/09/2017
- BIJKER, Wiebe E; HUGHES, Thomas P; PINCH, Trevor; DOUGLAS, Deborah G. **The social construction of technological systems**: New directions in the sociology and history of technology Massachusetts: MIT Press, 1987.
- BOCZKOWSKI, Pablo J. **Digitizing the news**: Innovation in online newspapers. Cambridge: MIT Press, 2004.
- _____. **News at work**. Chicago: The University of Chicago, 2010.
- BOCZKOWSKI, Pablo J. ANDERSON, C. W. **Remaking the News: Essays on the Future of Journalism Scholarship in the Digital Age**. MIT Press, 2017.
- BOCZKOWSKI, Pablo; LIEVROUW, Leah A. Bridging STS and communication studies: Scholarship on media and information technologies. In: HACKETT, Edward J.,

AMSTERDAMSKA, Olga, LYNCH, Michael, WAJCMAN, Judy. **The handbook of science and technology studies**, v. 3, The MIT Press, 2008, p. 949-977.

BROWN, James J. **Ethical programs: Hospitality and the rhetorics of software**. University of Michigan Press, 2015.

BUEGER, Christian; STOCKBRUEGGER, Jan. Actor-Network Theory: Objects and Actants, Networks and Narratives. In: MCCARTHY, Daniel. (Ed.) **Technology and World Politics: An Introduction**. Melbourne: Routledge, 2017. Disponível em: <www.researchgate.net/publication/281375658>. Acesso em: 13 jul. 2016.

CALLON, Michel. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay. IN: LAW, John. **Power, Action and Belief: a New Sociology of Knowledge?** London: Routledge and Kegan Paul, 1986. p. 196-233.

_____. Society in the making: the study of technology as a tool for sociological analysis. IN: BIJKER, Wiebe E; HUGHES, Thomas P; PINCH, Trevor; DOUGLAS, Deborah G. **The social construction of technological systems: New directions in the sociology and history of technology**. Massachusetts: MIT Press, 1987. p. 83-103.

CALLON, Michel; LASCOUMES, Pierre; BARTHE, Yannick. **Acting in an uncertain world: An essay on Technical Democracy**. Traduzido por Graham Burchell. MIT press, 2009.

CARLSON, Matt. The Robotic Reporter. **Digital Journalism**, 2014. DOI: 10.1080/21670811.2014.976412

_____. **Journalistic Authority: Legitimizing News in the Digital Era**. Columbia University Press, 2017.

CARR, David. **The Robots Are Coming! Oh, They're Here**. The New York Times 19/10/2009 Disponível em: <https://mediadecoder.blogs.nytimes.com/2009/10/19/the-robots-are-coming-oh-theyre-here/> Acessado em: 06/09/2017

CARREIRA, Krishma Anaísa Coura. **Notícias Automatizadas: A evolução que levou o jornalismo a ser feito por não humanos**. 2017. Dissertação (Mestrado em Comunicação Social) Universidade Metodista de São Paulo, São Paulo.

CLERWALL, Christer. Enter the Robot Journalist, **Journalism Practice**, 8:5, 519-531, 2014. DOI: 10.1080/17512786.2014.883116

CNBC Names Narrative Science to 2015 Disruptor 50 List. Business Wire. Disponível em: <http://www.businesswire.com/news/home/20150512005325/en/CNBC-Names-Narrative-Science-2015-Disruptor-50#.VX-tBqYVIFU> | Publicado em: 12/05/2015 | Acessado em 15/05/2015.

COHEN, Sarah. HAMILTON, James T. TURNER, Fred. Computational journalism. **Communications of the ACM** v. 54, n. 10, p. 66-71, 2011. DOI: 10.1145/2001269.1001288

DAWSON, Ross. **The rise of robot journalists**. Blog pessoal 15/04/2010 Disponível em: https://rossdawson.com/blog/the_rise_of_robot_journalists/ Acessado em: 07/09/2017

DIAKOPOULOS, Nicholas. **A functional roadmap for innovation in computational journalism**, 2011. Disponível em: http://www.nickdiakopoulos.com/wp-content/uploads/2007/05/CJ_Whitepaper_Diakopoulos.pdf Acessado em: 13/12/2017

_____. Algorithmic accountability: Journalistic investigation of computational power structures. **Digital Journalism** v. 3, n. 3, p. 398-415, 2014.

DÖRR, Konstantin Nicholas. Mapping the field of Algorithmic Journalism. **Digital Journalism**, 2015. DOI: 10.1080/21670811.2015.1096748

DÖRR, Konstantin Nicholas. HOLLNBUCHNER, Katharina. Ethical Challenges of Algorithmic Journalism, **Digital Journalism**, 2016. DOI: 10.1080/21670811.2016/1167612

EMPSON, Rip. **Attn Sports fans, journalists, and parents of Young athletes: GameChanger is on line two**. TechCrunch 15/06/2011 Disponível em: <https://techcrunch.com/2011/06/15/attn-sports-fans-journalists-and-parents-of-young-athletes-gamechanger-is-on-line-two/> Acessado em: 09/09/2017

FASSLER, Joe. **Can the computer at Narrative Science Replace Paid Writers?** The Atlantic 12/04/2012. Disponível em: <http://www.theatlantic.com/entertainment/archive/2012/04/can-the-computers-at-narrative-science-replace-paid-writers/255631/> - Acessado em: 22/05/2015.

FINLEY, Klint. **This news-writing bot is now free for everyone**. Wired. 20/10/2015 Disponível em: <https://www.wired.com/2015/10/this-news-writing-bot-is-now-free-for-everyone/> Acessado em: 30/10/2017

FISCHER, Frederick. **Interview with Ken Schwenke about Robot Journalism**. Vídeo (4m17s) 15/09/2014 Disponível em: <https://www.youtube.com/watch?v=J2Nq33G4cOo&t=2s> Acessado em: 12/10/2017

FRANKEL, Stuart. **Data Scientists Don't Scale**. Harvard Business Review. Disponível em: <http://hbr.org/2015/05/data-scientists-dont-scale> - Acessado em: 24/05/2015.

FREIRE, Letícia de Luna. Seguindo Bruno Latour: notas para uma antropologia simétrica. **Comum**, Rio de Janeiro, v.11, n.26, p. 46-65, 2006.

GANI, Aisha. HADDOU, Leila. **Could robots be the journalists of the future?** The Guardian 16/03/2014 Disponível em: <https://www.theguardian.com/media/shortcuts/2014/mar/16/could-robots-be-journalist-of-future> Acessado em: 15/01/2018

GILLESPIE, Tarleton. The relevance of Algorithms IN: GILLESPIE, Tarleton. BOCZKOWSKI, Pablo J. FOOT, Kirsten. **Media technologies: Essays on communication, materiality and society** MIT Press, 2014, p.167-194.

GLAHN, Harry R. Computer-produced worded forecasts. **Bulletin of the American Meteorological Society**, v. 51, n. 12, p. 1126-1131, 1970.

GOLDBERG, Eli. DRIEDGER, Norbert, KITTREDGE, Richard I. Using Natural-Language Processing to Produce Weather Forecasts. **IEEE Expert** 9 (2): 45-53, 1994.

GORDON, Rich. **Machine-Generated a Threat to Journalists? I think not**. 19/10/2009. Disponível em: <http://mediashift.org/2009/10/machine-generated-news-a-threat-to-journalists-i-think-not292/>. Acessado em 06/09/2017

GRAEFE, Andreas. **Guide to Automated Journalism**. Tow Center for Digital Journalism. Janeiro, 2016. Disponível em: <http://towcenter.org/research/guide-to-automated-journalism/>. Acessado em: 04/05/2016

GRAEFE, Andreas. HAIM, Mario. HAARMANN, Bastian. BROSIUS, Hans-Bernd. Readers' perception of computer-generated news: Credibility, expertise, and readability. **Journalism**, 2016. DOI: 10.1177/1464884916641269

GREENFIELD, Rebecca. **Robot Journalism Isn't Scary, It's Just Plain Bad**. The Atlantic. 20/03/2012. Disponível em: <https://www.theatlantic.com/technology/archive/2012/03/robot-journalism-isnt-scary-its-just-plain-bad/330324/> Acessado em: 12/09/2017

_____. **Robot Journalism Still Doesn't Sound So Scary**. The Atlantic. 25/04/2012. Disponível em: <https://www.theatlantic.com/technology/archive/2012/04/robot-journalism-still-doesnt-sound-scary/328893/> Acessado em 12/09/2017

GREINER, Andrew. **Algorithm writes sports stories without human intervention**. NBC Chicago. 12/10/2009 Disponível em: <http://www.nbcchicago.com/news/local/Who-Needs-Reporters-When-Theres-Monkeys-64014622.html>. Acessado em 07/09/2017

GRIMM, Joe. **StatSheet Network Automates Hundreds of Sports Stories from Databases**. Poynter 23/11/2010 Disponível em: <http://www.poynter.org/2010/statsheet-network-automates-hundreds-of-sports-stories-from-databases/108463/> Acessado em 07/09/2017

GRUSIN, Richard. Da remediação à premediação: ou de como a sensação de imediatismo da sociedade digital dos anos 1990 evoluiu para um clima de contínua antecipação do futuro no século XXI. Revista **MATRIZES**, vol. 7, nº 2, 2013. Disponível em: <http://www.matrizes.usp.br/index.php/matrizes/article/view/504/pdf>. Acesso em 28/06/2016.

HAIM, Mario; GRAEFE, Andreas. Automated News: Better than expected?. **Digital Journalism**, v. 5, n. 8, p. 1044-1059, 2017.

HAMILTON, James T. TURNER, Fred. **Accountability through algorithm: Developing the field of computational journalism**. Report from the Center for Advanced Study in the Behavioral Sciences (CASBS). Stanford University, Summer workshop, 2009.

HOLANDA, André; LEMOS, André. Do Paradigma ao Cosmograma: Sete Contribuições da Teoria Ator-Rede para a Pesquisa em Comunicação. In: Encontro Anual da Associação Nacional dos Programas de Pós-Graduação em Comunicação, 22. 4-7 jun. 2013, Salvador. Disponível em: http://compos.org.br/data/biblioteca_2050.pdf. Acesso em: 23 mai. 2017.

KELLY, Cathal. **Can a computer cover a ball game?** The Toronto Star 23/10/2009. Disponível em: https://www.thestar.com/sports/baseball/2009/10/23/can_a_computer_cover_a_ball_game.html Acessado em: 10/10/2017

KLEIN, Scott. **How to edit 52,000 stories at once**. ProPublica 24/01/2013 Disponível em: <https://www.propublica.org/nerds/how-to-edit-52000-stories-at-once> Acessado em: 09/10/2017

KIRKLAND, Sam. **'Robot' to write 1 billion stories in 2014 – but will you know it when you see it?**. Poynter 21/03/2014. Disponível em: <http://www.poynter.org/news/media-innovation/244113/robot-to-write-1-million-stories-in-2014-but-will-you-know-it-when-you-see-it/> - Acessado em 22/05/2015.

LAFLEUR, Jennifer. **Methodology on Our Educational Opportunity Project: About the Data and Our Analysis** ProPublica 30/06/2011 Disponível em: <https://www.propublica.org/article/opportunity-gap-methodology> Acessado em: 09/10/2017

LAFRANCE, Adrienne. **A computer tried (and failed) to write this article**. The Atlantic 08/06/2016. Disponível em: <https://www.theatlantic.com/technology/archive/2016/06/story-by-a-human/485984/> Acessado em: 10/05/2017

LAGE, Nilson. **O lead clássico como base para automação do discurso informativo**. Curso de jornalismo da Universidade Federal de Santa Catarina. Florianópolis, 1997.

_____. **A era das máquinas inteligentes, Ano I**. Florianópolis, 2003.

LATOUR, Bruno. Where are the missing masses? Sociology of a few mundane artifacts. IN: BIJKER, Wiebe E. LAW, John (eds). **Shaping Technonology / Building Society: Studies in Sociotechnical Change**. Cambridge: MIT Press, 1992.

_____. **Jamais fomos modernos**. Rio de Janeiro: Editora 34, 1994^a.

_____. On technical mediation – philosophy, sociology, genealogy. **Common Knowledge**, v. 3, nº 2, p. 29-64, 1994^b.

_____. The Berlin key or how to do words with things. **Matter, materiality and modern culture**, p. 10-21, 2000^a.

_____. **Ciência em ação**. São Paulo: UNESP, 2000^b.

_____. **A Esperança de Pandora**: ensaios sobre a realidade dos estudos científicos. Bauru: Edusc, 2001.

_____. **Reagregando o social**: uma introdução à teoria do ator-rede. Salvador: Edufba, 2012.

_____. **An inquiry into modes of existence**. Harvard University Press, 2013.

_____. **Cogitamus**: Seis cartas sobre as humanidades científicas. Tradução de Jamille Pinheiro Dias. São Paulo: Editora 34, 2016.

LAW, John. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. **Systemic practice and action research**, v. 5, n. 4, p. 379-393, 1992.

_____. STS as method. In: FELT, Ulrike et al. (Ed.). **The handbook of science and technology studies**. MIT Press, 2016.

LECOMPTE, Celeste. **Automation in the Newsroom**: How algorithms are helping reporters expand coverage, engage audiences and respond to breaking news. NiemanReports 01/09/2015 Disponível em: <http://niemanreports.org/articles/automation-in-the-newsroom/> Acessado em: 02/10/2016.

LEOPOLD, Wendy. **J-Schools Can Play a Role in Bringing Innovation to News Industry**. Site da Northwestern University. 02/06/2009. Disponível em: http://www.northwestern.edu/newscenter/stories/2009/06/medill_innovation.html Acessado em: 09/10/2017

LEVENSON, Eric. **L.A. Times Journalist Explains How a Bot Wrote His Earthquake Story for Him.** The Wire 17/03/2014. Disponível em: <https://www.theatlantic.com/technology/archive/2014/03/earthquake-bot-los-angeles-times/359261/> Acessado em: 09/10/2017

LEVY, Steven. **Can an algorithm write a better news story than a human reporter?** Wired 24/04/2012 Disponível em: <http://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/> Acessado 31/05/2015.

LEWIS, Seth C. WESTLUND, Oscar. Big data and journalism: Epistemology, expertise, economics, and ethics. **Digital Journalism** v. 3, n. 3, p. 447-466, 2014.

LIEVROUW, L. Materiality and media in communication and technology studies: An unfinished project. In: GILLESPIE, Tarleton, BOCZKOWSKI, Pablo J., FOOT, Kirsten A. **Media technologies: Essays on communication, materiality and society** MIT Press, 2014, p. 21-51.

LIN II, Rong-Gong. **Revenge of Y2K? A software bug might have caused false alert for big (and very old) earthquake.** Los Angeles Times. 22/06/2017. Disponível em: <http://www.latimes.com/local/lanow/la-me-earthquakesa-earthquake-68-quake-strike-s-near-ista-vista-calif-jyhw-htmlstory.html> Acessado em: 24/10/2017

LINDEN, Carl-Gustav. Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?. **Digital Journalism**, v. 5, n. 2, p. 123-140, 2017.

LOHR, Steve. **In case you wondered, a real human wrote this column.** The New York Times. 10/09/2011 Disponível em: <http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html?mcubz=0> Acessado em: 07/09/2017

LOVE, Dylan. **If you don't think robots can replace journalists, check out this article written by a computer.** Business Insider Australia. 10/07/2014 Disponível em: <https://www.businessinsider.com.au/narrative-science-quill-gamechanger-2014-7> Acessado em: 15/01/2018

MANOVICH, Lev. **Software Takes Command.** Nova York: Bloomsbury, 2013.

MARCONDES, Pyr. **Mande seus jornalistas embora, que meu software resolve o resto para você.** Proxima 15/03/2017 Disponível em: <http://www.proxima.com.br/home/proxima/blog-do-pyr/2017/03/15/manda-seus-jornalistas-embora-que-meu-software-resolve-o-resto-para-voce.html> Acessado em 30/08/2017.

MARCONI, Francesco. SIEGMAN, Alex. **The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines.** Associated Press. 05/04/2017 Disponível em: https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf Acessado em: 18/10/2017

MARCONI, Francesco. **Study: News automation by AP increases trading in financial markets.** Associated Press. 08/12/2016 Disponível em: <https://insights.ap.org/industry-trends/study-news-automation-by-ap-increases-trading-in-financial-markets> Acessado em: 23/10/2017

MARSHALL, Sarah. **Robot reporters: A look at the computers writing the news.** Journalism.co.uk 12/03/2013 Disponível em: <https://www.journalism.co.uk/news/robot-reporters-how-computers-are-writing-la-times-articles/s2/a552359/> Acessado em: 08/09/2017

MCCORMICK MAGAZINE. **The Intersection of Technology and Journalism.** Fall 2011. Disponível em: <http://www.mccormick.northwestern.edu/magazine/fall-2011/the-intersection-of-technology-and-journalism.html> Acessado em: 08/10/2017

MEDILL MAGAZINE. **Letters to the Editor.** Winter 2009. Issue 74. Disponível em: http://mikemoyer.com/wp-content/uploads/2010/01/FINAL_MedillMag74.pdf Acessado em: 08/10/2017

MEYER, Robinson. **How a California Earthquake Becomes the News: An Extremely Precise Timeline.** The Atlantic 19/03/2014 Disponível em: <https://www.theatlantic.com/technology/archive/2014/03/how-a-california-earthquake-becomes-the-news-an-extremely-precise-timeline/284506/> Acessado em: 09/10/2017

MONNERAT, Alessandra. **Grupo de cientistas de dados trabalha no primeiro robô-jornalista do Brasil para reportar sobre projetos de leis na Câmara.** Blog Jornalismo nas Américas 12/01/2018 Disponível em: <https://knightcenter.utexas.edu/pt-br/blog/00-19181-grupo-de-cientistas-de-dados-trabalha-no-primeiro-robo-jornalista-do-brasil-para-repor> Acessado em: 06/02/2018

MONTAL, Tal; REICH, Zvi. I, Robot. You, Journalist. Who is the Author? Authorship, bylines and full disclosure in automated journalism. **Digital Journalism**, v. 5, n. 7, p. 829-849, 2017.

MOROZOV, Evgeny. **A robot stole my Pulitzer!** Slate 19/03/2012 Disponível em: http://www.slate.com/articles/technology/future_tense/2012/03/narrative_science_robot_journalists_customized_news_and_the_danger_to_civil_discourse.html Acessado em 31/05/2015.

_____. **Os robôs vão substituir os jornalistas?** Folha de S. Paulo 19/03/2012 Disponível em: <http://www1.folha.uol.com.br/columnas/evgenymorozov/2012/03/1062594-os-robos-vao-substituir-os-jornalistas.shtml> Acessado em 31/05/2015.

NARRATIVE SCIENCE INC., Chicago, IL, Nathan Nichols, Michael Justin Smathers, Lawrence Birnbaum, Kristian Hammond, Lawrence E. Adams. **Method and apparatus for triggering the automatic generation of narratives.** 13/186,308. US nº PI 8,775,161 B1. 19 jul. 2011 ; 8 jul. 2014. Disponível em: <http://www.google.com.br/patents/US8775161> Acesso em: 3 mar. 2017

NARRATIVE SCIENCE INC., Chicago, IL, Lawrence A. Birnbaum, Kristian Hammond, Nicholas D. Allen, John R. Templon. **System and Method for using data to automatically generate a narrative story.** 12/779,636. US nº PI 8,688,434 B1. 13 mai. 2010; 1 abr. 2014. Disponível em: <https://www.google.ch/patents/US8688434> Acesso em: 3 abr. 2017

NARRATIVE SCIENCE. **Case Study: Forbes.com.** 2013 Disponível em: <http://resources.narrativescience.com/h/i/83535927-case-study-forbes> Acessado em: 31/05/2015

NOVAK, Matt. **Robot Journalist Accidentally Reports on Earthquake from 1925.** Gizmodo. 22/06/2017. Disponível em: <https://gizmodo.com/robot-journalist-accidentally-reports-on-earthquake-fro-1796325030> Acessado em: 24/10/2017

NPR. **Program Creates Computer-Generated Sports Stories.** NPR All Things Considered. 10/01/2010. Disponível em: <http://www.npr.org/templates/story/story.php?storyId=122424166> Acessado em: 06/09/2017

ON the media. Apresentado por Brooke Gladstone. **Murder Ink**. Rádio WNYC 16/02/2007
Entrevista com Jill Leovy. Disponível em: <http://www.wnyc.org/story/129146-murder-ink>
Acessado em: 08/09/2017

OREMUS, Will. **The first News report on the L.A. Earthquake was written by a robot**. Slate 17/03/2014 Disponível em: http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html - Acessado em: 09/10/2017

PARTRIDGE, Eric. **A short etymological dictionary of modern English**. Routledge and Kegan Paul, 1958.

PEREDA, Cristina F. **El periodismo se enfrenta al reto de los robots que elaboran noticias**. El País 21/07/2014. Disponível em: http://sociedad.elpais.com/sociedad/2014/07/11/actualidad/1405101512_992473.html - Acessado em: 24/05/2015

PETERS, Christopher; BROERSMA, Marcel. **Rethinking Journalism Again**. 2016.

PINCH, Trevor; LEUENBERGER, Christine. **Studying scientific controversy from the STS perspective**. Paper presented at the EASTS Conference "Science Controversies and Democracy, 2006.

PODOLNY, Shelley. **If an algorithm wrote this, how would you even Know?**. New York Times 07/03/2015 Disponível em: <http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html> - Acessado em: 22/05/2015

REID, Alastair. **How Homicide Report tells the 'true story' of LA's violent crime**. Journalism.co.uk 28/01/2014 Disponível em: <https://www.journalism.co.uk/news/how-the-homicide-report-tells-the-true-story-of-la-s-violent-crime/s2/a555713/> Acessado em: 08/09/2017

REITER, Ehud; DALE, Robert. Building applied natural language generation systems. **Natural Language Engineering**, v. 3, n. 01, p. 57-87, 1997.

REITER, Ehud. DALE, Robert. **Building Natural Language Generation Systems**. Cambridge: Cambridge University Press, 2000.

REITER, Ehud. SRIPADA, Somayajulu. HUNTER, Jim. YU, Jim. DAVY, Ian. Choosing Words in Computer-Generated Weather Forecasts. **Artificial Intelligence** 167: 137–69. 2005

RUTKIN, Aviva. **Rise of robot reporters: when software writes the news**. New Scientist. 21/03/2014. Disponível em: <https://www.newscientist.com/article/dn25273-rise-of-robot-reporters-when-software-writes-the-news/> Acessado em: 09/10/2017

SÁ, Nelson de. **CIA financia projeto de análise de dados**. Folha de S. Paulo 16/06/2013. Disponível em: <http://www1.folha.uol.com.br/fsp/mundo/114251-cia-financia-projeto-de-analise-de-dados.shtml> - Acessado em: 31/05/2015

SANTOS, Márcio Carneiro dos. Textos gerados por software – surge um novo gênero jornalístico? **Revista Estudos da Comunicação**. Curitiba, v.15. n.38, p. 274-291. Set/dez 2014.

_____. Narrativas automatizadas e a geração de textos jornalísticos: a estrutura de organização do *lead* traduzida em código. **Brazilian Journalism Research**. Volume I. Número I. 2016. P. 160-185.

SCHMIDT, Samantha. **A massive earthquake was reported in California Wednesday – by mistake**. The Washington Post. 22/06/2017. Disponível em: <https://www.washingtonpost.com/news/morning-mix/wp/2017/06/22/a-massive-earth-quake-was-just-reported-in-california-turns-out-it-happened-in-1925/> Acessado em: 24/10/2017

SCHONFELD, Erick. **Automated news comes to sports coverage via StatSheet**. Tech Crunch 12/11/2010 Disponível em: <https://techcrunch.com/2010/11/12/automated-news-sports-statsheet/> Acessado em: 07/09/2017

SCHWENCKE, Ken. **Earthquake: 4.4 quake strikes Los Angeles; no damage, injuries reported** Los Angeles Times 17/03/2014 Disponível em: <http://www.latimes.com/tn-gnp-earthquake-47-strikes-in-los-angeles-early-monday-20140317-story.html> Acessado em: 09/10/2017

SIMONITE, Tom. **Robot Journalist Finds New Work on Wall Street**. MIT Technology Review, 09/01/2015. Disponível em: <http://www.technologyreview.com/news/533976/robot-journalist-finds-new-work-on-wall-street/> Acessado em: 15/06/2015

SISMONDO, Sergio. **An introduction to Science and technology studies**. Oxford: Blackwell Publishing, 2010.

SRIPADA, Somayajulu; REITER, Ehud; DAVY, Ian. SumTime-Mousam: Configurable marine weather forecast generator. **Expert Update**, v. 6, n. 3, p. 4-10, 2003.

STROSS, Randall. **When the software is the Sportswriter**. The New York Times 27/11/2010 Disponível em: <http://www.nytimes.com/2010/11/28/business/28digi.html?mcubz=0> Acessado em: 07/09/2017

SUNDAR, S. Shyam. Exploring receivers' criteria for perception of print and online news. **Journalism & Mass Communication Quarterly**, v. 76, n. 2, p. 373-386, 1999.

THURMAN, Neil; DÖRR, Konstantin; KUNERT, Jessica. When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences. **Digital Journalism**, v. 5, n. 10, p. 1240-1259, 2017.

TSOTSIS, Alexia. **Former Crunchies Finalist StatSheet Receives 1.3 million in Series A**. Tech Crunch 04/08/2010 Disponível em: <https://techcrunch.com/2010/08/04/former-crunchies-finalist-statsheet-recieves-1-3-million-in-series-a/> Acessado em: 07/09/2017

TURING, Alan. Computing intelligence and machinery. **Mind**, v. 59, n. 2236, 1950.

TURNER, Fred. Actor-Networking the News. **Social Epistemology**. Vol. 19 No. 4. October-December 2005 p. 321-324

ULANOFF, Lance. **Need to write 5 million stories a week? Robot reporters to the rescue**. Mashable 01/07/2014 Disponível em: http://mashable.com/2014/07/01/robot-reporters-add-data-to-the-five-ws/#r4J2zh_3Ogqh Acessado em: 15/01/2018

VAN DALEN, Arjen. The Algorithms behind the headlines, **Journalism Practice**, 6:5-6, p. 648-658, 2012. DOI: 10.1080/17512786.2012.667268

VAN DER KAA, Hille; KRAHMER, Emiel. Journalist versus news consumer: The perceived credibility of machine written news. In: **Proceedings of the Computation+ Journalism Conference**, Columbia University, New York. 2014.

VENTURINI, Tommaso. Diving in magma: How to explore controversies with actor-network theory. **Public understanding of science**, v. 19, n. 3, p. 258-273, 2010.

_____. Building on faults: how to represent controversies with digital methods. **Public understanding of science**, v. 21, n. 7, p. 796-812, 2012.

VOETS, Johan. **Data and robots make newspapers and journalists disappear**. Numrush. 06/02/2013 Disponível em: <http://numrush.com/2013/02/06/data-and-robots-make-newspapers-and-journalists-disappear/> Acessado em: 15/01/2018

WALKER, Rob. **Who's afraid of robot journalists?** Yahoo. 09/07/2014. Disponível em: <https://finance.yahoo.com/news/whos-afraid-of-robot-journalists-91182873564.html> Acessado em: 09/10/2017

WINICK, Erin. Bridging the communications gap between human and machine. MIT Technology Review 15/03/2018 Disponível em: <https://www.technologyreview.com/s/610512/bridging-the-communications-gap-between-human-and-machine/> Acessado em: 30/03/2018

WINNER, Langdon. **Autonomous Technology**: Technics-out-of-control as a theme in political thought. MIT Press, 1977.

YOUNG, Mary Lynn; HERMIDA, Alfred. From Mr. and Mrs. outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. **Digital Journalism**, v. 3, n. 3, p. 381-397, 2015.