



**FACULDADE DE TECNOLOGIA E CIÊNCIAS SOCIAIS APLICADAS – FATECS**  
**CURSO: JORNALISMO**

VINÍCIUS DE SOUSA MENDONÇA  
21496375

**NOTÍCIAS GERADAS POR SOFTWARE: O JORNALISMO SEM REPÓRTER**

**BRASÍLIA**  
**2016**

VINÍCIUS DE SOUSA MENDONÇA

**NOTÍCIAS GERADAS POR SOFTWARE: O JORNALISMO SEM REPÓRTER**

Monografia apresentada como requisito para a conclusão do curso de Jornalismo pela Faculdade de Tecnologia e Ciências Sociais Aplicadas do Centro Universitário de Brasília – UniCEUB.

Orientador: Prof. Vivaldo Sousa

**BRASÍLIA**

**2016**

VINÍCIUS DE SOUSA MENDONÇA

**NOTÍCIAS GERADAS POR SOFTWARE: O JORNALISMO SEM REPÓRTER**

Monografia apresentada como requisito para a conclusão do curso de Jornalismo pela Faculdade de Tecnologia e Ciências Sociais Aplicadas do Centro Universitário de Brasília – UniCEUB.

Orientador: Prof. Vivaldo Sousa

Brasília, 20 de junho de 2016

---

Prof. Vivaldo Sousa  
Orientador

---

Prof. Luiz Cláudio Ferreira  
Examinador

---

Prof.<sup>a</sup> Renata Carvalho  
Examinadora

## AGRADECIMENTO

*Agradeço em primeiro lugar a Deus, sem o qual eu nada posso realizar. Sou grato também aos meus pais, pelo apoio que sempre me prestaram, e ao orientador, prof. Vivaldo Sousa, pelo auxílio oferecido na elaboração deste trabalho.*

## RESUMO

O objetivo deste trabalho é apresentar uma reflexão sobre as recentes aplicações de algoritmos de automação textual em redações jornalísticas. Por meio de revisão bibliográfica, é apresentada uma contextualização histórica sobre o uso de técnicas computacionais no jornalismo, desde seus primórdios, até a utilização de soluções tecnológicas das ciências da computação na produção de conteúdo informacional. Também é apresentado um resumo sobre o surgimento e o avanço histórico das técnicas de Inteligência Artificial, até o desenvolvimento de ferramentas de geração automática de textos em linguagem natural. Por fim, são apontadas previsões de autores para o desenvolvimento futuro do uso de softwares de automação no jornalismo.

Palavras-chave: Jornalismo Computacional. Algoritmos. Automação. Inteligência Artificial.

## **ABSTRACT**

The aim of this work is to present a reflection about the recent applications of textual automation algorithms in newsrooms. Through literature review, a historical contextualization is presented on the use of computational techniques in journalism, from its beginnings, until the use of technological solutions of computer science in the production of informational content. It's also presented an overview of the emergence and the historic breakthrough of Artificial Intelligence techniques, up to the development of tools that automatically generate texts in natural language. Finally, some theoretical predictions for the future development of the use of automation software in journalism are pointed.

Keywords: Computational Journalism. Algorithms. Automation. Artificial intelligence.

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	<b>07</b>
<b>1. REFERENCIAL TEÓRICO</b> .....	<b>09</b>
<b>2. REVOLUÇÃO DOS DADOS</b> .....	<b>12</b>
<b>3. JORNALISMO COMPUTACIONAL</b> .....	<b>15</b>
<b>4. INTELIGÊNCIA ARTIFICIAL E COMUNICAÇÃO</b> .....	<b>19</b>
<b>5. AUTOMAÇÃO DO TEXTO JORNALÍSTICO</b> .....	<b>24</b>
5.1. FUNDAMENTOS.....	26
5.2. POSSIBILIDADES.....	27
5.3. DESAFIOS.....	29
<b>6. CENÁRIOS PARA O FUTURO</b> .....	<b>31</b>
<b>CONCLUSÕES PRELIMINARES</b> .....	<b>34</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>36</b>

## INTRODUÇÃO

No início do capítulo no qual discute a tecnologia, no livro *Por que estudar a mídia?*, Roger Silverstone (2005, p.45) declara: “Não iremos longe em nosso interesse pela mídia sem inquirir a tecnologia”. De fato, não é tarefa difícil observar como os meios de comunicação caminham lado a lado com os meios tecnológicos. O que conhecemos como mídia hoje é, entre outros fatores, fruto de um processo de assimilação de invenções e aparatos técnicos nas diversas fases que envolvem a atividade da comunicação. Esse caminho histórico passa por invenções técnicas como a imprensa de Gutenberg, a fotografia, o telégrafo, o rádio, o telefone, a televisão, o computador, a internet, a telefonia móvel, entre tantas outras (STAVELIN, 2013).

A história do uso de computadores como ferramenta aplicada para auxiliar repórteres a extrair informação relevante a partir de dados não é tão recente, tendo se originado há mais de meio século. Em 1952, a rede de TV americana CBS utilizou pela primeira vez um computador para calcular, a partir dos resultados parciais, o vencedor das eleições presidenciais nos Estados Unidos (DEFLEUR, 1997; COX, 2000).

De lá para cá, diversas iniciativas e práticas que colocam a computação como auxiliar nas redações têm sido empreendidas. Uma das novidades mais recentes é a aplicação de técnicas de inteligência artificial na transformação de dados estruturados em texto jornalístico. Graefe (2016, p.9, tradução nossa) define a automação no jornalismo como o “processo de usar software ou algoritmos para gerar notícias automaticamente, sem a intervenção humana”.

O uso de softwares para a produção de conteúdo jornalístico já é realidade em locais como Estados Unidos e países europeus, onde grandes empresas de mídia têm empregado algoritmos em suas redações, próprios ou produzidos por companhias especializadas na geração automatizada de textos. Alguns exemplos são os boletins econômicos sobre o lucro trimestral de empresas que a agência de notícias *Associated Press* publica sem intervenção humana, com o auxílio do software *Wordsmith*, da empresa *Automated Insights*, e a geração automática de



textos sobre a ocorrência de terremotos que o *Los Angeles Times* realiza através de algoritmo criado por um de seus colaboradores (GRAEFE, 2016).

O presente trabalho tem o objetivo de apresentar um resumo sobre o processo histórico que tornou possível o advento de técnicas de automação computacional na construção do texto jornalístico e organizar uma reflexão sobre o tema, com base nas constatações, previsões e discussões levantadas por autores, de diversas partes do mundo, sobre as transformações que a área jornalística tem enfrentado e ainda pode enfrentar nos próximos anos com a expansão da aplicação de algoritmos e elementos de inteligência artificial como ferramentas autônomas para a produção textual.

## 1. REFERENCIAL TEÓRICO

Ao refletir sobre como abordar o tema da automação computacional nas redações jornalísticas em um trabalho científico, evidenciou-se a possibilidade de organizar uma reflexão sobre o assunto em língua portuguesa com um resgate resumido do panorama histórico que direcionou ao estado atual e trazendo à luz as discussões que têm sido propostas por autores de diversas partes do mundo. Por ser uma atividade que tem sido empregada na área da comunicação muito recentemente, seus efeitos ainda não atingiram a produção noticiosa brasileira de forma concreta, e por isso as discussões ainda se encontram em fases iniciais.

No entanto, como esses artifícios computacionais têm avançado cada vez mais em outros países, a produção teórica nesses locais também tem sido maior e possibilita uma observação um pouco mais aprofundada sobre as impressões que autores têm apresentado sobre o assunto.

A pesquisa e organização do corpo bibliográfico reunido para o referencial teórico desse trabalho se deu através da busca de artigos, livros e demais produções que se relacionassem com a temática em diversas fontes científicas disponíveis na internet. Para tanto foram utilizadas uma série de palavras-chave relacionadas ao tema, como jornalismo robô, *automated journalism* e *artificial intelligence journalism*, em buscas que incluíram tanto sites especializados em reunir trabalhos científicos, quanto resultados obtidos através de buscadores da web.

Também foram incluídas na pesquisa obras que, apesar de não serem direcionadas à discussão direta do tema que aqui se coloca, foram fundamentais para a apreensão do desenvolvimento histórico do uso de ferramentas computacionais na comunicação, possibilitando a descrição resumida que se apresenta nesse trabalho dos importantes passos que, ao longo do tempo, têm aproximado o jornalismo cada vez mais das ciências da computação.

Dentre os autores consultados na realização da pesquisa destaca-se Graefe (2016), que faz uma introdução ao tema da automação nas redações jornalísticas e resume o atual status da aplicação da técnica, além de elucidar brevemente sobre o funcionamento básico dos algoritmos e apresentar potenciais e limitações. O autor

também levanta as principais questões e implicações da automação para jornalistas, consumidores de notícias, organizações midiáticas e para a sociedade.

Dörr (2015) procura responder ao questionamento sobre a possibilidade de softwares baseados nos sistemas de geração de linguagem natural (Natural Language Generators – NLG) serem capazes de realizar funções que tradicionalmente sempre foram praticadas por jornalistas. Para isso, o autor realiza um mapeamento das empresas que oferecem tecnologia de NLG na área jornalística e conduz entrevistas com representantes de cada uma dessas companhias.

Stavelin (2013) apresenta uma ampla discussão sobre a interseção entre as práticas jornalísticas e as tecnológicas das ciências da computação. O autor procura identificar confusões causadas pela diversidade de termos que se referem a atividades próximas, mas que nem sempre tratam dos mesmos assuntos e busca apontar para as potencialidades que a tecnologia tem apresentado ao ramo do jornalismo.

Cox (2000) descreve um panorama histórico do uso de computadores como auxiliares nas redações jornalísticas desde 1952, quando um computador foi utilizado para calcular as probabilidades no resultado das eleições nos Estados Unidos, até técnicas mais recentes de Reportagens Assistidas por Computador (CAR, na sigla em inglês).

Van Dalen (2012) analisa quais podem ser as consequências da automatização de certas atividades jornalísticas, em especial as de caráter rotineiro. O autor realiza uma pesquisa com base em textos publicados por jornalistas na internet como reação às iniciativas de grandes empresas de mídia em automatizar seções da produção noticiosa. Uma das conclusões é que a maioria dos avaliados tem reconhecido a automação como oportunidade de auxílio nas atividades mais básicas, liberando mais tempo para ser gasto com aquelas que realmente exigem as habilidades de um jornalista humano.

Latar (2015) levanta uma série de questionamentos sobre as implicações do advento da produção jornalística automatizada. O autor discute quais as possibilidades podem ser alcançadas em um futuro que cada vez mais providenciará

as bases para essas tecnologias, através da expansão do registro de informações sobre as mais diversas características da sociedade.

Anderson, Bell e Shirky (2012) realizam um detalhado trabalho de análise sobre a atual situação do jornalismo, demonstrando como mudanças provocadas por diversos fatores são inevitáveis e apontando as direções para a adaptação necessária para a sobrevivência no ramo.

Clerwall (2014) apresenta os resultados de um experimento que buscou elucidar a percepção de leitores sobre textos automatizados, em comparação com a produção de jornalistas humanos. As conclusões demonstraram que os participantes da pesquisa não foram capazes de diferenciar qual texto foi automatizado e qual foi produzido por um jornalista. Além disso, atribuíram a sensação de maior credibilidade e objetividade à produção do software, apesar de acharem a leitura desse conteúdo menos prazerosa.

Corrêa e Bertocchi (2012) promovem uma discussão sobre as transformações que a área jornalística atravessa, principalmente com relação ao advento da web 3.0, ou web semântica, valorizando o papel de curadoria que o jornalismo desempenha diante da explosão informacional.

Normande (2013) propõe uma discussão inicial sobre o tema, em língua portuguesa, apresentando as iniciativas que começavam a aparecer em maior número naquele ano. De acordo com a autora, elementos apresentados no Paradigma do Jornalismo Digital em Base de Dados – JDBD, proposto por Susana Barbosa, se relacionam com a aplicação de ferramentas de automação na narrativa jornalística.

Os citados autores serviram como principal fonte para a construção do presente trabalho. Apesar de se tratar de tema recente, em decorrência das possíveis implicações relacionadas ao advento da utilização de técnicas de automação nas redações, o assunto tem sido cada vez mais frequente nas publicações científicas, principalmente de pesquisadores voltados para o relacionamento entre tecnologia e comunicação.

## 2. REVOLUÇÃO DOS DADOS

O acelerado desenvolvimento tecnológico alcançado nas últimas décadas tem desencadeado mudanças fundamentais na sociedade humana. De acordo com Silverstone (2005, p.45), vivenciamos atualmente a “era da informação”:

[...] estamos no meio de uma revolução tecnológica [...] de amplas consequências, revolução na geração e disseminação da informação. Novas tecnologias, novas mídias, cada vez mais convergentes pelo mecanismo da digitalização, estão transformando o tempo e o espaço sociais e culturais. Esse novo mundo não para: 24 horas de noticiário, 24 horas de serviços financeiros.

Esse é o cenário que se revela. A era da informação é a era da digitalização, na qual, cada vez mais, os complexos processos do mundo real são traduzidos e armazenados em forma de dados. Entre os fatores que dão possibilidade a essa revolução tecnológica está o crescente volume e variedade de informações produzidas sobre as mais diversas atividades humanas e da natureza, viabilizado pela presença cada vez mais ampla de dispositivos digitais móveis, sensores que captam informações sem parar, armazenamento de dados cada vez mais baratos. Há ainda o rápido desenvolvimento e a acelerada difusão das capacidades de processamento computacional, algoritmos, surgimento de processos que permitem o aprendizado por máquinas e processamento de dados em grande escala (MAYER-SCHÖNBERGER e CUKIER, 2013; LEWIS, 2015).

Essa ampliação continuada da capacidade de acumulação de dados codificados digitalmente tem viabilizado o armazenamento de uma diversidade cada vez maior de tipos de informações, como relata Fidalgo (2007, p.155):

A recolha e a cumulação de dados, referentes a todos os aspectos da vida humana, do que existe e do que se faz, são uma das características mais marcantes da sociedade actual, também conhecida justamente por sociedade da informação. As superfícies, os volumes, as temperaturas, as velocidades, são indicadores de quantificação de tudo o que existe. Desde a órbita dos planetas à estrutura dos átomos, do macro ao micro, inventaria-se progressivamente tudo o que é sólido, líquido e gasoso. E o mesmo se diga do que se faz, o que se compra nos supermercados, o combustível que se mete no carro, as idas ao médico, o que se receita, os consumos de água e de energia, as taxas de natalidade e de mortalidade, as sondagens de opinião e os índices de confiança.

Nesse contexto, uma das constatações que se tornam evidentes é a dificuldade humana de lidar com uma quantidade tão elevada de informações em tempo útil e com esforço razoável. Toda essa sobrecarga de dados disponíveis através da rede mundial de computadores dificulta um acompanhamento manual para extrair dessa massa informacional o que há de relevante. De acordo com Corrêa e Bertocchi (2012, p.131):

Constata-se uma abundância informativa e uma disseminação de múltiplas narrativas sobre os acontecimentos do mundo, criadas a partir de distintas fontes, e apresentadas em formatos para todos os gostos. Contudo, inexistente a garantia de transformação de dados em informação e tampouco estes em conhecimento por parte do homem.

Essa revolução tecnológica traz consigo, inevitavelmente, transformações nos processos de qualquer área que se relacione com informação, entre elas, obviamente, o jornalismo. Com uma quantidade de informações cada vez maior à disposição, não é possível manter as mesmas atividades e rotinas tradicionais de uma era analógica. Segundo Corrêa e Bertocchi (2012, p.131):

Estamos diante de um cenário cibercultural para o mundo informativo que disponibiliza para as redações contemporâneas uma rede mundial de computadores com um excesso de informações nunca visto [...]. Tal cenário é, ao mesmo tempo, previsto por diferentes analistas [...] como o futuro do jornalismo e também como uma nova onda de mudanças paradigmáticas.

Há, no entanto, uma vasta gama de possibilidades que se desvela para a área da comunicação, permitindo realizações inéditas que não poderiam se concretizar em um mundo analógico. De acordo com Stavelin (2013, p.4, tradução nossa):

A mudança de um fluxo de trabalho analógico para o digital coloca a computação como componente central na produção de notícias. Isso permite maior variedade para os usuários finais, automação de tarefas monótonas para as redações, e possibilita ao jornalista manusear e analisar o crescente conteúdo em bancos de dados gigantescos que são de relevância para os cidadãos. Para o jornalismo, a computação é um caminho promissor para a produção de notícias, especialmente para os que são capazes de lidar com computadores de acordo com necessidades específicas, usando a programação como um método jornalístico.

A revolução dos dados, no entanto, não se limita ao aumento vertiginoso da quantidade de informação disponível. Além disso, a velocidade desse processo se

dá de forma cada vez mais acelerada. Mayer-Schönberger e Cukier (2013) relatam que a explosão na quantidade de dados se deu primeiramente em ciências como a astronomia e genética, nos anos 2000, quando foi cunhado o termo “big-data”, e agora este conceito tem se expandido para diversas outras áreas da atividade humana. De acordo com eles, o conceito inicial de big-data estava relacionado à incapacidade técnica de lidar com uma quantidade tão grande de informações. No entanto, o significado do termo se expandiu, como definem:

Big-data refere-se a coisas que alguém é capaz de fazer em larga escala que não poderia ser concretizado em uma escala menor, como extrair insights ou criar novas formas de valor, de maneiras que alteram mercados, organizações, e o relacionamento entre cidadão e governos etc. Mas isso é só o começo. A era do big data desafia a forma como vivemos e interagimos com o mundo. (MAYER-SCHÖNBERGER e CUKIER, 2013, p.6, tradução nossa).

Para a área da comunicação, em especial o jornalismo, essa revolução na quantidade de informações e na velocidade em que dados dos mais diversos tipos são gerados tem provocado consequências de diversas ordens. Lewis e Westlund (2015, p.2, tradução nossa) retratam quatro aspectos que são afetados por esse cenário, nos sentidos epistemológico, das habilidades, econômico e ético:

Enquanto o fenômeno big-data levanta várias questões relevantes para os meios noticiosos, alguns dos mais essenciais se relacionam com a legitimação das novas alegações sobre o conhecimento e a verdade (epistemologia); a negociação de *status* profissionais, autoridades, e conjunto de habilidades enquanto novas especializações são desenvolvidas e aplicadas (habilidades); o potencial e os desafios de novas capacidades, recursos, inovações, criação de valores e oportunidades de receita (econômico); e as apreensões levantadas por esses avanços com relação à normas e valores que guiam a tomada de decisão humana e o desenvolvimento de sistemas tecnológicos (ético).

As mudanças no jornalismo, alavancadas pelo advento da era digital, têm se dado em diversas áreas da atividade. Podemos citar desde transformações mais elementares, como o uso de processadores digitais de texto, até alterações de caráter estrutural como a utilização da internet como meio de suporte com o surgimento do webjornalismo, o uso de informações de bancos de dados para auxiliar ou compor histórias, até as recentes iniciativas de aplicação de algoritmos para a automação de tarefas nas redações.

### 3. JORNALISMO COMPUTACIONAL

Apesar do desenvolvimento de tecnologias digitais ser um fato consideravelmente recente na história da humanidade, a relação entre jornalismo e computação teve início há mais de sessenta anos. Na noite do dia 4 de novembro de 1952, a rede de televisão americana *CBS* utilizou três computadores *Remington Rand UNIVAC* para calcular o resultado das eleições presidenciais a partir dos resultados preliminares na contagem de votos. Todas as pesquisas indicavam uma disputa apertada entre Dwight Eisenhower e Adlai Stevenson. Com 7% dos votos contabilizados, os cálculos do computador indicaram uma vitória esmagadora de Eisenhower. Como a previsão era contrária a todas as expectativas, os editores do programa televisivo decidiram não divulgar e fingiram que havia um problema técnico com a máquina. Quando a conclusão dos votos foi divulgada, acabaram revelando que o prognóstico do computador havia sido extremamente preciso, se aproximando bastante do resultado final (DEFLEUR 1997; COX, 2000).

De acordo com Cox (2000), após essa primeira utilização de computadores com fins jornalísticos, a inserção da informática nas redações americanas se deu em primeiro lugar nas áreas administrativas, para fins de contabilidade e cálculos financeiros, durante a década de 1960. “Na década de 1970, jornais começaram a usar computadores para a produção também, o que levou à fase da informação” (COX, 2000, p.6, tradução nossa).

Um dos marcos na utilização de computadores para fins jornalísticos foi a cobertura de uma série de revoltas populares na cidade americana de Detroit, em 1967, conduzida por Philip Meyer, um importante nome no desenvolvimento das técnicas de reportagem assistida por computador (*computer-assisted reporting* – CAR). Meyer conduziu uma pesquisa entre afro-americanos durante os motins e utilizou um computador *IBM 360* para analisar os dados. De acordo com Cox (2000, p.7, tradução nossa):

A análise revelou que, ao contrário da hipótese da maioria, pessoas que haviam frequentado a faculdade tinham a mesma probabilidade de participar das revoltas como os que haviam abandonado os estudos. A história rendeu a Meyer um prêmio Pulitzer e marcou o início da nova era em reportagem assistida por computador.



O objetivo principal de Meyer era aplicar o uso de ferramentas de pesquisa das ciências sociais ao jornalismo. Como havia estudado também uma linguagem de programação para computadores IBM, o jornalista uniu conhecimentos nas duas áreas para realizar os cálculos da pesquisa de forma computadorizada.

Um ano após a publicação de Meyer, o jornal *The Miami Herald* contratou estudantes de direito da Faculdade de Miami para inserirem informações sobre decisões judiciais da região em um computador, com o objetivo de encontrar correlações nos dados. Os resultados foram publicados e a iniciativa se tornou a primeira utilização de computadores para analisar dados governamentais (COX, 2000).

Enquanto isso, Meyer continuava progredindo na utilização de estatísticas das ciências sociais em busca de informações que pudessem ser divulgadas. Em 1973, publica o livro intitulado *Precision Journalism*, no qual lança o conceito de que jornalismo e ciência são áreas similares. Segundo Cox (2000), ao se tornar científico o jornalismo tinha agora que lidar com a questão de como tratar dados. A autora aponta algumas considerações de Meyer:

Meyer identificou seis passos a serem seguidos ao se trabalhar com dados. O primeiro é a coleta. Em segundo lugar, é necessário armazenar esses dados, de preferência em computadores. É preciso ser capaz de recuperar os dados guardados, não apenas os próprios, mas também os armazenados por outros. Os jornalistas devem ser capazes de analisar os dados e procurar por padrões. Com uma grande quantidade de dados, o jornalista precisa ser capaz de reduzi-los. Finalmente, o jornalista precisa saber como comunicar os dados. (COX, 2000, p.8, tradução nossa).

Durante o restante dos anos 70, repórteres de outros grandes jornais americanos começaram a fazer uso do auxílio de computadores para a investigação de histórias. Cox (2000) relata iniciativas, durante esse período, de jornais como o *The New York Times* e o *The Philadelphia Inquirer*.

Nos anos 80, iniciou-se nos EUA o processo de popularização dos microcomputadores. De acordo com Cox (2000), num primeiro momento os repórteres começaram a usar seus próprios equipamentos, que eram utilizados primordialmente para processamento de texto, substituindo as máquinas de escrever. Depois, as empresas de mídia passaram a fornecer computadores para as

redações, com propósitos como de permitir a conexão a bancos de dados online. Segundo Cox (2000, p.10, tradução nossa):

A partir dos anos 80, as redações começam a utilizar bancos de dados de diversas maneiras. Primeiro, começaram a guardar recortes de notícias antigas. Depois, a utilizar os bancos de dados para informações de segundo plano de suas próprias histórias. Algumas redações também desenvolveram bancos de dados para tópicos específicos. Como complemento, usavam computadores para analisar registros governamentais.

Seguindo os passos de pioneiros na utilização de computadores como ferramenta jornalística, diversas técnicas foram sendo desenvolvidas ao longo dos anos seguintes, com descrições nas quais é comum se intercalarem algumas características, mas que carregam também detalhes próprios. Entre elas foram cunhados termos, em inglês, como *precision journalism*, *computer-assisted reporting (CAR)*, *data-driven journalism*, *data journalism*, *database journalism* e *computational journalism*. Stavelin (2013, p.46, tradução nossa) esclarece as semelhanças e diferenças entre as técnicas:

Todas têm em comum a abordagem computacional do jornalismo como marca dessa atividade; também separam os praticantes dessas técnicas dos jornalistas “convencionais”. Todas requerem habilidades especiais no uso mais avançado de computadores. [...] Mas, há diferenças. Em suma, jornalismo de precisão (*precision journalism*) enfatiza o uso de métodos científicos, CAR (*computer-assisted reporting*) enfatiza o uso de ferramentas digitais, jornalismo de banco de dados (*database journalism*) prima pela estrutura do armazenamento e recuperação de informações, jornalismo de dados e dirigido a dados (*data journalism* e *data-driven journalism*) enfatizam a procura de histórias em conjuntos de dados, enquanto jornalismo computacional (*computational journalism*) foca a união de valores da computação e do jornalismo na criação de ferramentas e aplicação de métodos. (Parênteses nossos)

Para os objetivos deste trabalho, nos interessa primordialmente o desenvolvimento e as características do chamado jornalismo computacional. De acordo com Stavelin (2013, p.49, tradução nossa), esse novo ramo teórico também tem pontos de contato com as outras técnicas apontadas, mas os ultrapassa em algumas características:

Jornalismo computacional está intercalado com todas essas áreas, mas também cobre outras fora da tradição centrada na história, do jornalismo realizado com computadores. Jornalismo computacional também é iniciado de fora para dentro das redações e é descrito como uma interseção entre jornalismo e ciência da computação.

Ou seja, as técnicas aplicadas no jornalismo computacional estão menos ligadas à resolução de questionamentos encontrados em histórias isoladas, como os usos tradicionais dos computadores como auxiliares em reportagens específicas, mas se expandem para a construção de ferramentas que objetivem solucionar problemas mais gerais, se aplicando a atividades mais amplas do jornalismo.

É nesse contexto teórico do jornalismo computacional que se inserem as técnicas de automação jornalísticas, com ferramentas que, em diversas ocasiões, são apresentadas de fora para dentro das redações, como no caso de empresas especializadas que oferecem serviço para os veículos, e que não se limitam a histórias isoladas, mas se apresentam como soluções replicáveis para um número ilimitado de produtos em áreas diversas.

#### 4. INTELIGÊNCIA ARTIFICIAL E COMUNICAÇÃO

“Podem as máquinas pensar?”. Essa pergunta formulada pelo matemático e filósofo inglês Alan Turing, em 1950, tem, desde então, sido foco de inúmeras discussões e indagações. Há quem se oponha, alegando serem as qualidades do pensamento inteligente restritas aos humanos. E há quem defenda que, se o cérebro não for mais que um conjunto de neurônios, organizados para executarem funções com base em interações elétricas e químicas, não haverá impedimento para o desenvolvimento de entidades máquinicas baseadas em construções lógicas que simulem redes neuronais e possam mimetizar o pensamento e apresentar inteligência semelhante à humana (HODGES, 1999; COPPIN, 2010).

O trabalho de Turing formou as bases para os estudos e discussões atuais relacionados à Inteligência Artificial (IA), termo cunhado após a morte do matemático, por John McCarthy, em 1956 (COPPIN, 2010, p.8). Turing desenvolveu, em um dos primeiros trabalhos sobre o assunto, o artigo *Computing Machinery and Intelligence*, publicado em 1950, as regras para um teste que, segundo ele, seria capaz de definir se uma máquina apresenta inteligência. O teste, denominado de “Jogo da Imitação”, consiste, em suma, na aferição da capacidade de uma máquina de se passar por um ser humano, após uma série de perguntas realizadas por um interrogador inconsciente sobre a natureza do interrogado (HODGES, 1999; GUNKEL, 2012).

As tentativas de se criar ferramentas automatizadas capazes de ter sucesso no teste de Turing não demoraram muito a ser empreendidas. A primeira robô, ou *chatbot*, como passaram a ser chamados esses programas, construída com o objetivo de tentar mimetizar a comunicação humana, batizada de Eliza, foi desenvolvida entre 1964 e 1966, por Joseph Weizenbaum, pesquisador do Instituto de Tecnologia de Massachusetts (PRIMO e COELHO, 2002; FERREIRA, 2008).

Como embrião dos algoritmos de conversação, o programa Eliza era simples, contava com apenas 204 linhas de código. O objetivo era simular as respostas de uma psicóloga a um cliente, que em muitas ocasiões utilizava as informações inseridas pelo usuário e respondia através de novas perguntas, o que criava a impressão de interesse e diálogo. As reações positivas do público chegaram a

surpreender o próprio Weizenbaum, sendo que até mesmo “muitos psiquiatras acreditaram de imediato que Eliza poderia ser desenvolvida até um sistema quase totalmente automatizado de psicoterapia” (PRIMO e COELHO, 2002, p.3).

Desde a criação do algoritmo que “dava vida” a Eliza, diversos outros programas foram desenvolvidos a partir do conceito fundamental no qual um usuário insere uma informação e recebe de volta uma resposta que se relacione com o *input*. Esses robôs têm sido usados em várias áreas, como educação e entretenimento. Ferreira (2008, p.4) apresenta uma lista de tipos de *bots* de acordo com o uso em que são aplicados:

Hoje, a aplicação dos chatterbots é variada, indo desde simples “psicólogos virtuais” até usos mais práticos para a indústria, como os FAQs (Frequently Asked Questions, “tutoriais” que ajudam os usuários com problemas), podendo ser divididos em: *chatterbots*, *news bots*, *fun bots*, *academic bots*, *bot design*, *commerce bots*, *government bots*, *knowledge bots*, *search bots*, *shopping bots*, *stock bots* e *update bots*, dentre outras categorias menores ou extremamente híbridas.

No entanto, embora tenham se desenvolvido, como apresenta Ferreira (2008) no estudo em que analisa quatro casos de *chatterbots* mais recentes, ainda não foi apresentado um sistema com complexidade suficiente para vencer o teste de Turing, “ainda que prêmios em dinheiro sejam regularmente oferecidos ao inventor do primeiro programa de computação que o faça” (COPPIN, 2010, p.8).

Coppin (2010, p.9) relata que “desde os anos 50, uma grande dose do otimismo original com a Inteligência Artificial foi perdida, tendo sido substituída por uma dose de realismo”. Cada vez mais as pesquisas em IA têm deixado em segundo plano as tentativas de alcançar a simulação do intelecto humano e se dedicado ao desenvolvimento de soluções para problemas mais práticos, usando como base a forma como o cérebro humano funciona (COPPIN, 2010).

É fundamental, neste contexto, observar que essa discussão, de raízes mais filosóficas que práticas, sobre a possibilidade de máquinas se igualarem aos humanos em sua capacidade cognitiva, se restringe a uma vertente específica nos estudos de Inteligência Artificial. Dentro desse campo há distinção entre IA forte e IA fraca, como destaca Coppin (2010, p.5):

Os seguidores da IA forte acreditam que, dispondo de um computador com suficiente capacidade de processamento e fornecendo a ele suficiente inteligência, pode-se criar um computador que possa literalmente pensar e ser consciente do mesmo modo que um humano é consciente. [...] Por outro lado, IA fraca é simplesmente a visão de que comportamento inteligente pode ser modelado e utilizado por computadores para solucionar problemas complexos. Este ponto de vista defende que apenas o fato de um computador agir inteligentemente não prova que ele seja verdadeiramente inteligente no sentido humano.

De fato, uma das definições de Inteligência artificial apresentadas por Coppin (2010, p.4) é a de que “Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos”. É instrutivo notar como essa definição é baseada na finalidade do uso da IA. Para ser capaz de solucionar problemas com níveis de complexidade diferenciados, técnicas que buscam referência nos modos mentais de seres biológicos, como animais e o homem, são empregadas. O objetivo não é criar seres conscientes, mas métodos capazes de responder às situações de forma similar aos seres com inteligência complexa.

Soluções originadas dos estudos em IA, criadas para darem respostas a questões práticas, estão atualmente em uso e se desenvolvem em várias áreas tecnológicas com as quais nos relacionamos. Ferreira (2008), por exemplo, destaca que o conceito dos *chatterbots* ultrapassa as tentativas de simulação de seres com vida própria, alcançando até mesmo ferramentas como o motor de buscas do *Google*, “que é nada mais que um site que usa determinada tecnologia para processar uma entrada (digitada no campo de busca do usuário) e devolver uma resposta (na forma de links de sites) [...]” (FERREIRA, 2008, p.6).

Quando se trata do uso de recursos de IA fraca, não há que se negar a ampla presença desses artifícios em diversas atividades humanas, em especial no campo da comunicação, como apresenta Gunkel (2012, p.2, tradução nossa):

Mesmo que alguém duvide da possibilidade de se alcançar o que tradicionalmente tem sido chamado de “inteligência artificial forte”, o fato é que nosso mundo já é povoado por artefatos semi-inteligentes ou dispositivos inteligentes que de modo crescente desempenham o papel não de meios de comunicação, mas de fontes ou receptores de informação.

Gunkel (2012) afirma que a comunicação é fundamental tanto para as teorias como para as práticas relacionadas aos estudos de Inteligência Artificial. Ele destaca que até mesmo o teste proposto por Turing tem por base a capacidade comunicativa para decidir se uma máquina é capaz de pensar (GUNKEL, 2012, p.5). O argumento do autor é que os estudos em comunicação precisam se adaptar ao cenário onde computadores não são apenas um meio que conecta pessoas (comunicação mediada por computador), mas tomam parte como agentes ativos nos processos de comunicação. Para Gunkel (2012, p.9, tradução nossa)

Ainda que haja debates nos círculos relacionados à inteligência artificial se o sucesso na simulação da comunicação em nível humano implica na presença de “inteligência” ou não, o que não é tema de debate é que as máquinas são de fato capazes de se comunicar com usuários humanos em uma variedade de contextos e de formas que, com frequência, não é possível distinguir da comunicação de uma outra pessoa. E para as pesquisas em estudos da comunicação – onde comunicação e não inteligência é o foco – isso faz toda a diferença.

É importante aqui destacar essa constatação de Gunkel (2012), sobre a capacidade de as máquinas realizarem trocas comunicativas com seres humanos, chegando ao ponto de não deixarem margem para que um leitor comum faça distinção entre o que é produto do intelecto humano e o que é resultado de processos automatizados, programados para transmitir informações em linguagem compreensível.

Essa possibilidade, até certo ponto alcançada, se deve ao desenvolvimento dos estudos de processamento de linguagem natural, ramo das pesquisas em Inteligência Artificial. De acordo com Coppin (2010), as primeiras tentativas de utilização dessas técnicas estavam relacionadas à busca de ferramentas que permitissem automatizar a tradução de textos entre idiomas diferentes. O autor relata que a tarefa não foi tão simples de ser realizada, como alguns acreditavam, principalmente em decorrência do grande nível de ambiguidades que as palavras costumam carregar: “Tradução por máquina não é possível usando simplesmente análise sintática e análise léxica: um conhecimento do mundo que estiver sendo discutido será também essencial, de modo a desambiguar o texto que está sendo traduzido” (COPPIN, 2010, p.515).

Essa área de pesquisa possibilitou o posterior surgimento de sistemas geradores de linguagem natural (*Natural Language Generators – NLG*, em inglês), com o objetivo de criar textos em linguagem corrente a partir do uso de dados. Já nos anos 70 começaram a ser utilizados esses sistemas para, por exemplo, geração automática de boletins climáticos (GRAEFE, 2016).

Os sistemas geradores de linguagem natural são as bases dos desenvolvimentos recentes de algoritmos voltados à automação, capazes de gerar textos inteligíveis a partir de bancos de dados para os mais diversos usos, como jogos eletrônicos, boletins econômicos, publicidade personalizada e também com fins jornalísticos.



## 5. AUTOMAÇÃO DO TEXTO JORNALÍSTICO

Após um terremoto atingir uma região do estado da Califórnia, nos Estados Unidos, o site do jornal *Los Angeles Times* foi o primeiro a noticiar o evento, em menos de 3 minutos após o abalo sísmico. Oremus (2014) reproduz o texto gerado pelo software (mantido em inglês para apresentar a produção original do algoritmo):

*A shallow magnitude 4.7 earthquake was reported Monday morning five miles from Westwood, California, according to the U.S. Geological Survey. The temblor occurred at 6:25 a.m. Pacific time at a depth of 5.0 miles. According to the USGS, the epicenter was six miles from Beverly Hills, California, seven miles from Universal City, California, seven miles from Santa Monica, California and 348 miles from Sacramento, California. In the past ten days, there have been no earthquakes magnitude 3.0 and greater centered nearby. This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.*

Esse feito foi possível graças a um algoritmo, escrito pelo jornalista e programador Ken Schwencke. Acordado pelo tremor de terra numa manhã de março de 2014, Schwencke foi direto ao computador para revisar o texto organizado pelo programa e publicá-lo. Nomeado de *Quakebot*, o algoritmo é fundamentado, de acordo com o próprio criador, em conceitos simples e funciona preenchendo lacunas em módulos textuais pré-estruturados com dados gerados e reportados pelo serviço de notificação de terremotos americano (GRAEFE, 2016).

Além do *Quakebot*, Schwencke organizou outros sistemas de automação para o *Los Angeles Times*. O *Homicide Report* é um desses projetos, criado com o objetivo de publicar de forma automática informações básicas dos homicídios na região de Los Angeles. Graefe (2016, p.16, tradução nossa) relata que “antes do lançamento do projeto, em janeiro de 2007, a versão impressa do *Times* cobria apenas dez por cento da média de mil assassinatos anuais praticados na área”.

Com o sistema implementado, todas as ocorrências passaram a ser relatadas, com base em dados estruturados fornecidos pela polícia local, contendo informações como data, local, horário, raça ou etnia, idade, jurisdição e vizinhança de todos os homicídios na região.

Ao apresentar a situação atual sobre a automação no jornalismo, Graefe (2016, p.5, tradução nossa) aponta que “grandes empresas de mídia, como *Associated Press, Forbes, The New York Times, Los Angeles Times* e *ProPublica* já começaram a automatizar conteúdo noticioso”. O autor ainda afirma que “apesar de a tecnologia estar nas fases iniciais, o jornalismo automatizado chegou nas redações e provavelmente para ficar” (GRAEFE, 2016, p.5, tradução nossa).

Em setembro de 2015, o site *Nieman Reports* publicou o artigo *Automation in the Newsroom: how algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news*, escrito pela jornalista Celeste Lecompte, no qual descreve o panorama do uso de ferramentas de automação nas redações jornalísticas, até aquele momento. Lecompte cita no artigo diversas iniciativas de empresas de mídia no sentido de automatizar tanto a produção de conteúdo jornalístico, como também outras etapas das rotinas que se dão nas redações, como análise de dados, a etiquetagem automática e o monitoramento de espaços de comentários e redes sociais (LECOMPTTE, 2015; LOKOT e DIAKOPOULOS, 2015).

Além de descrever projetos de automação de conteúdo que tiveram origem no interior das redações, como os já citados *Quakebot* e o *Homicide Report*, do *Los Angeles Times*, o artigo revela também iniciativas que têm o auxílio externo de empresas especializadas na tarefa de, a partir do uso de algoritmos especiais, transformar dados estruturados em textos que se aproximam da linguagem natural humana. Entre essas empresas está a americana *Automated Insights*, que no início de 2015 anunciou que seu software de automação, o *Wordsmith*, estava produzindo e publicando três mil artigos sobre lucros empresariais a cada trimestre, para a agência de notícias *Associated Press*. Antes da implementação do sistema, a agência estima que conseguia produzir, manualmente, cerca de trezentos artigos por trimestre (LECOMPT, 2015; LOKOT e DIAKOPOULOS, 2015; GRAEFE, 2016).

A *Automated Insights* não está sozinha no ramo. De acordo com Dörr (2015), em um mapeamento de empresas que fornecem algum tipo de serviço relacionado à geração de linguagem natural, entre treze companhias encontradas, onze ofereciam produto relacionado à automação textual para o jornalismo, em diversos países e

idiomas diferentes. Dörr (2015, p.12, tradução nossa) apresenta alguns nomes dentre as mais importantes:

Considerando o tamanho da empresa, clientes e produtos já lançados, a Automated Insights e Narrative Science são as mais importantes forças do Jornalismo de Algoritmos nos Estados Unidos. Na Alemanha são a Retresco e Aexea, na França Syllabs e na Grã Bretanha a Arria.

A concorrente da *Automated Insights* nos Estados Unidos, *Narrative Science*, fornece seus serviços de automação, desde 2012, para o site da *Forbes*, em um trabalho similar ao realizado pela *Associated Press*, de divulgação de lucros empresariais. O software de automação da empresa, *Quill*, também foi utilizado pela organização *ProPublica*, em 2013, no projeto batizado de *Opportunity Gap*, no qual foram utilizados dados governamentais para automatizar a geração de textos informativos numa plataforma que abarcou informações sobre mais de 52 mil escolas americanas (GRAEFE, 2016).

## 5.1. FUNDAMENTOS DA AUTOMAÇÃO

O ponto de partida para a criação dos algoritmos de automação jornalística são as técnicas de geração de linguagem natural (NLG), provenientes dos estudos de inteligência artificial aplicados à linguística. As etapas necessárias para a automação envolvem o acesso aos dados, que providenciam o conteúdo do texto, uma etapa intermediária na qual esses dados passam por um processamento com base em critérios previamente estabelecidos, e a parte final, quando o texto é produzido para ser publicado automaticamente ou para ser editado por um jornalista. De acordo com Dörr (2015, p.5, tradução nossa):

Os dados, “o ingrediente principal” dos geradores de linguagem natural no jornalismo, podem ser acessados através de APIs públicas ou privadas. Para a geração de textos automáticos ligados a esse conteúdo, consequentemente todo produto jornalístico baseado em NLGs, códigos individuais, regras e dicionário precisam ser codificados e adaptados. Sendo assim, sistemas NLG só funcionam com base em regras específicas pré-determinadas. Isso inclui o processo de criação linguística assim como as decisões sobre quais informações escondidas entre os dados devem ser processadas e transformadas em linguagem natural. [...] Isso envolve parâmetros como tamanho do texto, conteúdo/fatos, formas de apresentação jornalística, tema, tonalidade e o local e tempo da publicação.

O elemento principal na produção automatizada de textos é essa etapa intermediária do processo, quando operações estatísticas predefinidas são aplicadas às informações provenientes dos dados para definir a relevância e o valor-notícia de cada acontecimento identificado.

## 5.2 POSSIBILIDADES

Graefe (2016, p.5, tradução nossa) indica alguns elementos que defensores das técnicas de automação costumam apresentar como positivos na aplicação dos algoritmos nas redações jornalísticas:

Algoritmos são capazes de gerar notícias mais rápido, numa escala maior e com potencial de cometer menos erros que jornalistas humanos. Podem usar os mesmos dados para escrever histórias em diversos idiomas e por ângulos diferentes, personalizando para as preferências de cada leitor. Algoritmos têm ainda o potencial de gerar notícias por demanda, produzindo textos em resposta a questões dos usuários sobre os dados.

A publicação em menos de três minutos do terremoto na Califórnia, a partir do texto produzido pelo *Quakebot*, do *Los Angeles Times*, exemplifica as possibilidades colocadas pela automação quanto à velocidade. Tudo que o algoritmo precisa são dos dados para ser capaz de gerar as notícias instantaneamente. Van Dalen (2012) afirma que essa é uma habilidade com a qual os jornalistas humanos não podem competir. De acordo com Dörr (2015, p.8, tradução nossa), “velocidade de publicação é importante no jornalismo digital e o conteúdo online pode ser continuamente atualizado se (em tempo real) dados estiverem disponíveis, sendo possível ignorar rotinas de produção”.

Outra alegação dos defensores da automação é que os algoritmos podem ser capazes de contribuir para o aumento da exatidão no jornalismo. Segundo Graefe (2016, p. 18, tradução nossa), “algoritmos não se distraem ou ficam cansados, e – assumindo que foram programados corretamente e que os dados de base são exatos – não cometem erros, como de digitação, de cálculos incorretos ou omissão de fatos”. De acordo com o autor, a *Associated Press* relatou queda na quantidade de erros de 7% para 1% nos artigos automatizados, em comparação

com o período quando todo o trabalho era feito de forma manual, principalmente na redução de erros de digitação e trocas nos valores numéricos (GRAEFE, 2016).

Uma importante característica da automação é a expansão da quantidade de conteúdo produzido e publicado. Graefe (2016) cita o aumento no número de artigos financeiros fornecidos pela *Associated Press* e o incremento nos casos de assassinatos divulgados pelo *Homicide Reports*, do *Los Angeles Times*. Ambos relataram um aumento de mais de dez vezes na produção de conteúdo após a implantação dos algoritmos. A automação abre espaço para que áreas que não recebiam atenção e não eram cobertas por falta de recursos, sejam também divulgadas. De acordo com Dörr, (2015, p.13, tradução nossa):

Porque a produção de notícias continua sendo um negócio que busca o lucro, é evidente que, por um lado, as empresas midiáticas procuram reduzir custos. Por outro, também estão desesperadas por novos produtos jornalísticos e maneiras de satisfazer a audiência. [...] Considerando que os custos com sistemas geradores de linguagem natural são baixos, comparados com jornalistas humanos, o jornalismo por algoritmos pode ser rentável.

Mais uma possibilidade ampliada pelo uso de algoritmos é a capacidade de aumento no nível de personalização na produção de conteúdo noticioso. Na medida em que cada vez mais os usuários dos meios online têm suas características pessoais disponibilizadas, as chamadas identidades digitais, a capacidade de moldar o conteúdo aos interesses e características de indivíduos ou comunidades se amplia (LATAR e NORDFORS, 2009). Para Graefe (2016, p.21, tradução nossa), a “automação permite prover informação que seja relevante para audiências muito pequenas. Em casos extremos é possível até criar notícias para uma pessoa”.

Em um estudo realizado em 2014, Chister Clerwall questionou sobre a percepção dos leitores de textos gerados por software em comparação com os tradicionais, escritos por jornalistas humanos. O pesquisador selecionou um grupo de quarenta e seis estudantes, que foram divididos em duas partes. Um grupo recebeu um texto gerado por algoritmo sobre resultados esportivos, alheios à fonte automatizada. O outro, um artigo com características semelhantes, mas escrito por um jornalista humano. Após a leitura, os participantes do estudo responderam a um formulário de perguntas sobre aspectos como qualidade e credibilidade do texto.

Os resultados apresentados demonstraram que o conteúdo gerado por software foi melhor avaliado que o produzido pelo jornalista em aspectos como ser descritivo, informativo, confiável e objetivo. Outro resultado interessante foi que, questionados sobre qual acreditavam ser a fonte do texto, os participantes não demonstraram capacidade clara de distinguir quando se tratava de produção humana ou automatizada (CLERWALL, 2014, p.11).

### 5.3 DESAFIOS

Para ser colocada em prática a automação no jornalismo apresenta algumas exigências, entre elas o acesso a dados estruturados de alta qualidade. De acordo com Graefe (2016), esse é um dos motivos que limitam a aplicação atual a áreas como finanças, esportes ou clima, pois essas categorias já apresentam um elevado grau de conversão em dados organizados. O autor cita o exemplo de uma tentativa anunciada pela *Associated Press* de automatizar notícias sobre eventos esportivos colegiais a partir de estatísticas divulgadas. O projeto enfrentou dificuldades devido à falta de qualidade dos dados, que “por serem inseridos pelos treinadores e não passarem por verificações, acabam desorganizados e com erros” (GRAEFE, 2016, p.22, tradução nossa).

Graefe (2016, p.5, tradução nossa) aponta que “o jornalismo automatizado é especialmente útil na geração de textos em tópicos rotineiros e repetitivos, para os quais estejam disponíveis dados limpos, corretos e estruturados”. Em assuntos mais complexos, que exigem habilidades mais apuradas, os programas de computador ainda são bastante deficientes, como apresenta o autor:

Assim que as respostas são validadas, os algoritmos podem contribuir com conhecimento. No entanto, essa contribuição se limita a responder questionamentos elaborados previamente, a partir da análise dos dados. Algoritmos não são capazes de usar o conhecimento para fazer novos questionamentos, detectar necessidades, reconhecer ameaças, resolver problemas, ou oferecer opiniões e interpretações em, por exemplo, assuntos que dizem respeito a mudanças sociais e políticas. (GRAEFE, 2016, p.23, tradução nossa).

Outro ponto fraco dos atuais algoritmos é a qualidade do texto, em comparação com a produção tradicional. Apesar de o estudo conduzido por Clerwall

(2014) ter indicado avaliações positivas quanto à confiabilidade e credibilidade, o conteúdo gerado por software ficou atrás do escrito pelo jornalista em aspectos como satisfação na leitura e qualidade do texto. De acordo com Graefe (2016, p.21, tradução nossa):

Os algoritmos atuais são limitados na compreensão e produção das nuances da linguagem humana, como humor, sarcasmo e metáforas. Notícias automatizadas podem se apresentar técnicas e entediantes e experimentos tem demonstrado que as pessoas preferem ler notícias escritas por humanos.

## 6. CENÁRIOS PARA O FUTURO

Uma das previsões mais otimistas sobre o futuro do uso de softwares geradores de notícias nas redações vem de alguém interessado na expansão desse mercado. Para um dos criadores e executivo da americana *Narrative Science*, Kristian Hammond, em cinco a dez anos 90% da produção jornalística deverá ser realizada por algoritmos (LATAR, 2015). Apesar do prognóstico bastante otimista, não é possível ignorar que, mesmo não alcançando a expectativa de Hammond, “é previsto que empresas de mídia passarão por mudanças dramáticas na próxima década, devido à introdução de processos automáticos de Inteligência Artificial em todos os aspectos da produção e disseminação de notícias” (LATAR, 2015, p.74, tradução nossa).

Van Dalen (2012, p.651, tradução nossa) apresenta dois fatores que indicam o motivo pelo qual o jornalismo automatizado é uma tendência que precisa ser levada a sério:

Primeiro, a profissão jornalística é cada vez mais comercial e se baseia nas lógicas de mercado. A automação do jornalismo se encaixa na tendência de busca por altos lucros a baixos custos. A segunda razão que pode fazer a automação ser bem sucedida é a qualidade das notícias com as quais está competindo. Artigos gerados por computador podem não ser capazes de competir com jornalismo de alta qualidade, oferecido por grandes empresas, que se atentam aos detalhes, análises, informações contextuais e apresentam uma linguagem mais viva ou humor. Mas para a informação que está disponível gratuitamente na internet o nível é relativamente mais baixo e conteúdo gerado automaticamente é capaz de competir.

O surgimento de artefatos tecnológicos capazes de provocar alterações no *status quo* sempre são objeto de discussões. Alguns defendem visões otimistas, outros apontam motivos para ter um olhar mais pessimista. Com o advento dos algoritmos nas redações não tem sido diferente (VAN DALEN, 2012). Entre os questionamentos que geralmente são levantados com o aparecimento de técnicas de automação estão as possibilidades de substituição de mão de obra humana, frente ao avanço das máquinas. Para Van Dalen (2012, p.653, tradução nossa), esse novo cenário pressiona os jornalistas a reconsiderarem seus papéis e habilidades mais fundamentais: “se a escrita por máquinas se tornar mais comum no futuro, jornalistas não poderão competir realizando apenas tarefas rotineiras, como artigos descritivos sobre esportes”.



Já os que preferem uma abordagem mais otimista acreditam que, com o avanço da tecnologia, os jornalistas terão mais tempo útil para realizar tarefas de maior complexidade, conforme as máquinas se ocupam das atividades básicas, repetitivas e de rotina. De acordo com Latar (2015, p. 67, tradução nossa):

Essa possível ameaça à profissão de jornalistas humanos é encarada por alguns otimistas meramente como mais uma ferramenta que vai liberá-los da necessidade de conduzir dispendiosas e, às vezes, perigosas investigações. O jornalista robô vai providenciar, assim esperam, um rascunho de uma notícia que eles irão editar e enriquecer com suas análises aprofundadas, suas perspectivas e talentos narrativos.

Graefe (2016, p.40, tradução nossa) concorda que deverão haver mudanças fundamentais na maneira como jornalistas realizam seu trabalho e também aponta para um futuro de união de forças entre algoritmos e jornalistas:

No futuro, jornalismo humano e automatizado deverão se tornar integrados e formar uma espécie de “casamento entre homem e máquina”. Jornalistas são aconselhados a focar em tarefas que algoritmos não são capazes de realizar, como análises em profundidade, entrevistas com pessoas-chave e reportagem investigativa. Enquanto a automação vai, provavelmente, substituir jornalistas que meramente cobrem tópicos rotineiros, a tecnologia também está gerando novos empregos no processo de desenvolvimento dos algoritmos geradores de notícias.

Essa união de esforços entre homens e máquinas levanta, ainda, uma série de questões éticas e de transparência. Alguns exemplos dizem respeito aos direitos autorais e à divulgação sobre até onde o trabalho é realização do jornalista ou do algoritmo (VAN DALEN, 2012). Segundo Graefe (2016, p.41, tradução nossa):

Pouco é conhecido sobre se os consumidores de notícias (como necessidade ou desejo) entendem como os algoritmos funcionam, ou sobre qual informação eles utilizam para gerar conteúdo. [...] Tais informações podem se tornar importantes em situações em que (a) erros sejam cometidos e (b) o conteúdo for personalizado de acordo com as necessidades e preferências do consumidor individual.

Essa possibilidade de personalização, citada por Graefe (2016), também é assunto de debates. Atualmente, já é prática comum que empresas como o *Google* utilizem informações coletadas sobre o cotidiano dos usuários, com o objetivo de oferecer conteúdo e publicidade que possam ser de interesse personalizado (GRAEFE, 2016). A questão que se apresenta é sobre quais seriam os efeitos de

oferecer informações jornalísticas utilizando essa mesma lógica de direcionamento a desejos e necessidades pessoais. Segundo Graefe (2016, p.39, tradução nossa):

Ao se fazer isso, consumidores de notícias diversos podem receber conteúdos noticiosos diferentes para as mesmas pesquisas digitadas, o que carrega o risco de uma parcial cegueira informacional, a hipótese chamada de “filtro bolha”. De acordo com essa ideia, a personalização levará os indivíduos a consumir mais e mais da mesma informação, conforme os algoritmos fornecem somente conteúdo que os usuários gostem de ler ou concordem. Conseqüentemente, as pessoas teriam menos probabilidade de serem confrontadas com informações que desafiem seus pontos de vista ou contradigam seus interesses, o que poderia ser arriscado para a formação da opinião pública em uma sociedade democrática.

Anderson, Bell e Shirky (2012, p.38, tradução nossa) apontam que uma mudança de paradigma importante nesse novo contexto é a necessidade de os jornalistas expandirem cada vez mais suas habilidades com relação a componentes tecnológicos que têm se tornado fundamentais nas redações:

Jornalistas precisam aprender a programar. É bem verdade que para ser fluente em linguagens de programação são necessárias habilidades muito desenvolvidas. Nem todo jornalista vai fazer isso, e nem todo jornalista deveria fazer isso. Mas, todo jornalista precisa entender em um nível básico o que é programação, o que ela poder fazer e como se comunicar com aqueles que são mais habilidosos nessa área.

Nesse sentido, há também que se discutir como as universidades e os cursos de jornalismo irão se adaptar ao novo cenário. Van Dalen (2012, p.656, tradução nossa) destaca que:

O aumento da automação do trabalho jornalístico tem importantes implicações para os educadores da área. Será tarefa deles providenciar aos jornalistas do futuro habilidades jornalísticas, técnicas e criativas. Essa mistura de habilidades será indispensável no mercado de trabalho do futuro, onde jornalistas competirão com algoritmos criando mais de 15 mil artigos por mês.

De acordo com Van Dalen (2012, p.655, tradução nossa) é importante notar que não serão somente os avanços tecnológicos sozinhos que irão ditar os rumos que serão trilhados a partir da expansão das técnicas de automação no jornalismo. Também contribuirão nesse processo “fatores sociais e o contexto institucional das companhias de mídia e do mercado, como a cultura de trabalho, posição das uniões de jornalistas e relacionamento entre empregados e patrões”.

## CONCLUSÕES PRELIMINARES

Buscamos apresentar neste trabalho, por meio de pesquisa bibliográfica, um panorama histórico da utilização de ferramentas tecnológicas no jornalismo, desde a introdução dos computadores como auxiliares na apuração e produção de conteúdo informacional, até as mais recentes aplicações de técnicas de automação e inteligência artificial na construção do texto jornalístico. Foi possível constatar que há mais de meio século a ciência da computação tem fornecido soluções para que jornalistas sejam capazes de trabalhar com dados, especialmente a partir da explosão na quantidade de informações disponíveis em formato digital que se tem observado nos últimos anos.

Nosso foco principal foi demonstrar como essa expansão na disponibilidade de dados estruturados tem transformado em realidade nas redações jornalísticas a utilização de algoritmos capazes de produzir textos noticiosos prontos para serem publicados automaticamente ou serem editados por um jornalista. Também buscamos apresentar as visões de pesquisadores com relação aos efeitos, possibilidades e desafios que acompanham o advento dessas técnicas.

A utilização de ferramentas de inteligência artificial aplicadas à construção de conteúdo jornalístico ainda está nos seus primórdios. Por isso, é difícil prever concretamente quais serão os próximos passos no desenvolvimento e uso dessas técnicas nas redações. No entanto, a partir do que se tem observado onde a automação já é aplicada e das constatações de teóricos, é possível enumerar algumas tendências:

- Em primeiro lugar, somente áreas nas quais estiverem disponíveis dados estruturados de alta qualidade poderão ser automatizadas;
- Os algoritmos poderão realizar atividades rotineiras, liberando tempo para os jornalistas se dedicarem a tarefas mais complexas e que exigem esforço e habilidades humanas;
- Jornalistas precisarão cada vez mais unir conhecimentos e habilidades da comunicação com os relacionados à área da ciência da computação. Deverão, por exemplo, adquirir conhecimentos relacionados à programação

de computadores, pelo menos em nível básico, que lhes permita trabalhar em colaboração com profissionais da computação;

- As universidades e os cursos de jornalismo necessitarão se adaptar ao novo cenário, ajustando suas matrizes curriculares para oferecer conhecimentos relacionados a essas novas habilidades que serão requisitos para a prática do jornalismo computacional;
- Deverá haver aumento na quantidade de informação jornalística disponível, com a possibilidade de ampliação na personalização do conteúdo para cada indivíduo ou comunidade;
- A expansão na produção de conteúdo e na cobertura de temas antes ignorados precisará ser vantagem suficiente para ultrapassar a deficiência na qualidade dos textos, quando comparados com a produção de jornalistas;
- Jornalistas deverão continuar como os principais produtores de conteúdo informacional de caráter mais elaborado e complexo, mas terão cada vez mais o auxílio de ferramentas tecnológicas para acessarem as informações necessárias para a construção de histórias.

Novas pesquisas na área deverão ser realizadas nos próximos anos, conforme a utilização da automação no jornalismo avance. Ainda são passíveis de maior esclarecimento questões como a receptividade dos consumidores de notícias à produção automatizada e quais deverão ser as regras relacionadas à ética e transparência na utilização dos dados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, Chris W.; BELL, Emily; SHIRKY, Clay. Post-industrial journalism: Adapting to the present. Columbia Journalism School, [Nova York], *online*, 2012. Disponível em: <<http://towcenter.org/research/post-industrial-journalism-adapting-to-the-present-2/>>. Acesso em: 9 de abr. 2016.
- CLERWALL, Christer. Enter the Robot Journalist: Users' perceptions of automated content. *Journalism Practice*, [S.l.], v. 8, n. 5, p. 519-531, 2014. Disponível em: <[http://www.academia.edu/11569516/Enter\\_the\\_robot\\_journalist](http://www.academia.edu/11569516/Enter_the_robot_journalist)>. Acesso em: 9 de abr. 2016.
- COPPIN, Ben. *Inteligência artificial*. Rio de Janeiro: LTC, 2010. p.668.
- CORRÊA, Elizabeth Saad; BERTOCCHI, Daniela. A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. *Matrizes*, São Paulo, v. 5, n. 2, 2012. p. 123-144. Disponível em: <<http://www.matrizes.usp.br/index.php/matrizes/article/viewFile/340/pdf>>. Acesso em: 9 de abr. 2016.
- COX, Melisma. The development of computer-assisted reporting. In: Association for Education in Journalism and Mass Communication, Southeast Colloquium, Chapel Hill, *Anais*, 2000. p. 1-22. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>>. Acesso em: 12 de mai. 2016.
- DEFLEUR, Margaret. *Computer-assisted investigative reporting: development and methodology*. Mahwah: Lawrence Erlbaum Associates, 1997. p.264.
- DÖRR, Konstantin Nicholas. Mapping the field of Algorithmic Journalism. *Digital Journalism*, [Nova York], p. 1-23, 2015. Disponível em: <[http://www.mediachange.ch/media//pdf/publications/MAPPING\\_THE\\_FIELD\\_OF\\_ALGORITHMIC\\_JOURNALISM\\_DoerrK\\_.pdf](http://www.mediachange.ch/media//pdf/publications/MAPPING_THE_FIELD_OF_ALGORITHMIC_JOURNALISM_DoerrK_.pdf)>. Acesso em: 9 de abr. 2016.
- FERREIRA, Leonardo. Trocas Comunicativas em Ambientes Virtuais e Inteligências Artificiais. In: Intercom - X Congresso de Ciências da Comunicação na Região Nordeste, *Anais*, São Luís, 2008. p.15. Disponível em: <<http://www.intercom.org.br/papers/regionais/nordeste2008/resumos/R12-0355-1.pdf>>. Acesso em: 9 de abr. 2016.
- FIDALGO, António. Data Mining e um novo jornalismo de investigação. In: BARBOSA, Suzana (Org.). *Jornalismo Digital de Terceira Geração*. Covilhã: Labcom, 2007. p. 155-168. Disponível em: <[http://www.livroslabcom.ubi.pt/pdfs/20110824-barbosa\\_suzana\\_jornalismo\\_digital\\_terceira\\_geracao.pdf](http://www.livroslabcom.ubi.pt/pdfs/20110824-barbosa_suzana_jornalismo_digital_terceira_geracao.pdf)>. Acesso em: 12 de abr. 2016.

GRAEFE, Andreas. Guide to Automated Journalism. Tow Center for Digital Journalism, [Nova York], *online*, 2016. Disponível em: <<http://towcenter.org/research/guide-to-automated-journalism/>>. Acesso em: 9 de abr. 2016.

GUNKEL, David. Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+1*, [S.l.], v. 1, n. 1, p. 1-25, 2012. Disponível em: <<http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1007&context=cpo>>. Acesso em: 9 de abr. 2016.

HODGES, Andrew. *Turing: um filósofo da natureza*. São Paulo: Unesp, 1999. p.66.

LATAR, Noam Lemelshtrich; NORDFORS, David. Digital Identities and Journalism Content: How Artificial Intelligence and Journalism May Co-Develop and Why Society Should Care. *Innovation Journalism*, [S.l.], v. 6, n. 7, p.1-47, 2009. Disponível em: <<http://www.innovationjournalism.org/archive/INJO-6-7.pdf>>. Acesso em: 9 de abr. 2016.

LATAR, Noam Lemelshtrich. The Robot Journalist in the Age of Social Physics: The End of Human Journalism?. In: ENAV, Gali. *The New World of Transitioned Media*. [S.l.]: Springer International Publishing, 2015. p. 65-80. Disponível em: <[https://d27ojnwysu5c5p.cloudfront.net/he/ha-sha-a\\_haben-tchumit/ha-sha-a\\_haben-tchumit/bt\\_gil-ve-noam-lemelshtrich-latar.pdf](https://d27ojnwysu5c5p.cloudfront.net/he/ha-sha-a_haben-tchumit/ha-sha-a_haben-tchumit/bt_gil-ve-noam-lemelshtrich-latar.pdf)>. Acesso em: 21 de mai. 2016.

LECOMPTE, Celeste. Automation in the Newsroom: How algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news. Nieman Reports, [S.l.], *online*, 2015. Disponível em: <<http://niemanreports.org/articles/automation-in-the-newsroom/>>. Acesso em: 12 de mai. 2016.

LEWIS, Seth; WESTLUND, Oscar. Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, [Nova York], v. 3, n. 3, p. 447-466, 2015. Disponível em: <<https://apps.cla.umn.edu/directory/items/publication/324155.pdf>>. Acesso em: 9 de abr. 2016.

LEWIS, Seth. Journalism in an Era of Big Data: Cases, Concepts, and Critiques. *Digital Journalism*, [Nova York], v. 3, n. 3, p. 321-330, 2015. Disponível em: <<https://apps.cla.umn.edu/directory/items/publication/324153.pdf>>. Acesso em: 9 de abr. 2016.

LOKOT, Tetyana; DIAKOPOULOS, Nicholas. News Bots: Automating news and information dissemination on Twitter. *Digital Journalism*, [Nova York], p. 1-18, 2015. Disponível em: <[http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/newsbots\\_final.pdf](http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/newsbots_final.pdf)>. Acesso em: 9 de abr. 2016.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big data: A revolution that will transform how we live, work, and think*. Londres: John Murray, 2013. p.256.

NORMANDE, Naara. A automatização da narrativa jornalística. *Estudos em Comunicação*, Salvador, nº 13, p.363-378, 2013. Disponível em: <<http://www.ec.ubi.pt/ec/13/pdf/EC13-2013Junho-14.pdf>>. Acesso em: 9 de abr. 2016.

OREMUS, Will. The First News Report on the L.A. Earthquake Was Written by a Robot. Slate, [S.l.], *online*, 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](http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html)>. Acesso em: 12 de mai. 2016.

PRIMO, Alex; COELHO, Luciano. Comunicação e Inteligência Artificial: interagindo com a robô de conversação Cybelle. In: MOTTA, Luiz et al. (Eds.). *Estratégias e culturas da comunicação ed.Brasília*. Brasília: Editora Universidade de Brasília, 2002. p. 83-106. Disponível em: <<http://www.ufrgs.br/limc/PDFs/cybelle.pdf>>. Acesso em: 9 de abr. 2016.

SILVERSTONE, Roger. Por que estudar a mídia. 2. ed. São Paulo: Loyola, 2005. p.302.

STAVELIN, Eirik. *Computational Journalism: When journalism meets programming*. 2013. 107 f. Tese (Doutorado em Filosofia) - Department of Information Science and Media Studies, University of Bergen, Bergen. 2013. Disponível em: <<http://bora.uib.no/bitstream/handle/1956/7926/dr-thesis-2013-Eirik-Stavelin.pdf>>. Acesso em: 9 de abr. 2016.

VAN DALEN, Arjen. The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, [S.l.], v. 6, n. 5-6, p. 648-658, 2012.