

**NOTÍCIAS
AUTOMATIZADAS,
GERAÇÃO DE
LINGUAGEM NATURAL E
A LÓGICA DO BOM
SUFICIENTE**

AUTOMATED JOURNALISM,
NATURAL LANGUAGE
GENERATION AND
THE LOGIC OF GOOD ENOUGH

PERIODISMO AUTOMATIZADO,
GENERACIÓN DE LENGUAJE
NATURAL Y LA LÓGICA DE LO
SUFICIENTEMENTE BUENO

**Krishma Carreira¹
Sebastião Squirra^{2, 3}**

RESUMO

A explosão de dados digitais e o avanço do campo da Inteligência Artificial permitiram o surgimento de algoritmos capazes de redigir e distribuir automaticamente notícias jornalísticas com rapidez e de forma customizada. A automatização da produção de narrativas entrega textos básicos, objetivos e sem sofisticação, formatando a função de jornalistas e contribui para a

¹ Jornalista, mestre pelo Programa de Pós-Graduação em Comunicação Social da Umesp de São Paulo. Pesquisadora integrante do ComTec - Grupo de Pesquisa em Comunicação e Tecnologias Digitais (www.comtec.pro.br). E-mail: krishmacarreira@gmail.com.

² Pesquisador com Doutorado pela ECA/USP, atualmente é professor do Programa de Pós-graduação em Comunicação da Umesp de São Paulo, onde lidera o ComTec - Grupo de Pesquisa em Comunicação e Tecnologias Digitais (www.comtec.pro.br). Email: ssquirra@gmail.com.

³ Endereço de contato dos autores (por correio): Universidade Metodista de São Paulo, Faculdade de Comunicação Social, Programa de Pós Graduação Em Comunicação Social. Rua do Sacramento 230 - Edifício Capa, Rudge Ramos, CEP: 09735-460, São Bernardo do Campo, SP, Brasil.

ampliação da chamada "bolha de filtros" definida por Eli Pariser. Neste texto, procuramos demonstrar que ao adotar a automação, redações nos Estados Unidos, Europa e China passaram a usar a lógica do *bom suficiente*, demonstrada por Robert Capps. Ou seja, a automatização da elaboração de notícias pode estar sendo adotada unicamente por seu caráter "industrial" e pelo custo radicalmente reduzido, não priorizando a diversificação, a cultura e a qualidade humana dos relatos. As reflexões centram-se em pesquisa bibliográfica de caráter exploratório e interdisciplinar, a partir da abordagem da Teoria Ator-Rede.

PALAVRAS-CHAVE: Jornalismo automatizado; robôs e jornalismo; algoritmo jornalista; futuro do jornalismo.

ABSTRACT

The explosion of digital data and the advance of the field of Artificial Intelligence allowed the emergence of algorithms capable of automatically writing and distributing news stories quickly and in a customized way. The automation of narrative production delivers basic, objective and unsophisticated texts, shaping the role of journalists and contributes to the expansion of the so-called "filter bubble" defined by Eli Pariser. In this text, we try to demonstrate that in adopting automation, essays in the United States, Europe and China have come to use the logic of good enough, demonstrated by Robert Capps. That is to say, the automation of the elaboration of news can be adopted solely by its "industrial" character and by the radically reduced cost, not prioritizing the diversification, the culture and the human quality of the stories. The reflections center on bibliographic research of an exploratory and interdisciplinary character, based on the approach of the Theory-Actor Network.

KEYWORDS: Automated journalism; robots and journalism; journalism algorithm; future of journalism

RESUMEN

La explosión de datos digitales y el avance del campo de la inteligencia artificial han permitido el surgimiento de algoritmos capaces de escribir y distribuir automáticamente las noticias periodísticas con velocidad y de una manera personalizada. La automatización de la producción de narrativas entrega escritos básicos, objetivos y poco sofisticados, formateando la función de los periodistas y colabora con la ampliación de la llamada “burbuja de filtros”, definida por Eli Pariser. En este texto, buscamos demostrar que, mediante la adopción de la automatizaciones, las redacciones de los periódicos en Estados Unidos, Europa y China comenzaron a utilizar la lógica de que es “bueno lo suficiente”, demostrada por Robert Capps. Es decir, la automatización de la preparación de las noticias puede estar siendo adoptada únicamente por su carácter “industrial” y por el costo radicalmente reducido, no priorizando, de esta manera, la diversificación, la cultura y la calidad humana de los relatos. Las reflexiones se encuentran centradas en la búsqueda bibliográfica de carácter exploratorio e interdisciplinario con base en el abordaje de la Teoría Actor-Red.

PALABRAS CLAVE: Periodismo automatizado; robots y periodismo; algoritmo de periodista; futuro del periodismo.

Recebido em: 08.02.2017. Aceito em: 20.04.2017. Publicado em: 01.05.2017.

Introdução

Desde que as tecnologias digitais interconectáveis foram introduzidas, o campo do Jornalismo enfrenta mudanças nas formas de produzir, distribuir e compartilhar a informação. Tal realidade impacta profundamente as empresas jornalísticas, os jornalistas, o conteúdo e, por conseguinte, a própria sociedade. Isto, pois a partir da década de 1960, os computadores começaram a ser usados em jornais como ferramentas de suporte da produção, através do conceito de CAR (*Computer-Assisted Reporting* ou Reportagem Auxiliada por Computador), uma tradição que foi iniciada nos Estados Unidos como uma forma de encorajar a adoção de investigações jornalísticas com base em dados informatizados (PARASIE; DAGIRAL, 2012, p. 2). A partir do uso do computador como “assistente” do repórter, houve uma explosão de dados digitais gerados por variados tipos de registros, atividades e transações, organizando as informações em multiplicadas redes de computadores, constituindo os Bancos de Dados *online*. Estes passaram a ser acessados por meio de muitos dispositivos fixos e móveis, de redes sociais, de arquivos *online* públicos, governamentais, empresarias, educacionais, dos sensores químicos e atmosféricos, de câmeras de segurança e através de formatos diversos. Para Mayer-Schönberger e Cukier (2013), este *tsunami* de dados pode ser chamado de *datafication*, termo sem tradução para o português, mas que pode ser entendido como a transformação de tudo em dado - incluindo o que nunca foi pensado como informação - e sua tradução em algo quantificável.

Os dados aumentaram ainda mais a partir da Internet das Coisas, também conhecida pela sigla IoT do inglês *Internet of Things*, que representa “um olhar sobre as possibilidades de conexão e troca de informação entre os objetos comuns, que quando se conectam à internet, oferecem novas formas de

utilização” (SANTOS, 2016a, p. 186). O pesquisador Márcio Carneiro dos Santos (2016a, p. 186) explica que, com a IoT, “relógios, sensores, eletrodomésticos e outros itens quando conectados, potencialmente, podem nos oferecer informação em tempo real sobre o que está acontecendo ao nosso redor, mesmo quando estamos distantes deles. Compreende-se dado como uma unidade básica de informação antes da interpretação do sujeito (FLORIDI, 2000) e informação, como dado dotado de relevância e propósito (DRUCKER apud DAVENPORT, 1998, p. 19).

A armazenagem de grandes quantidades de dados de fontes diversas e o uso de ferramentas e processos mais complexos que os usados em bancos de dados comuns passaram a ser conhecidos como *Big Data*. Este conjunto de técnicas é empregado para encontrar valor e transformar os dados em informações confiáveis, relevantes, que vão ajudar na tomada de decisões e na geração de *insights*. Com a mudança da quantidade de dados, a essência deles também foi alterada e eles transformaram-se em fonte poderosa e de valor econômico destacável (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 13-14). Os autores ressaltam que os humanos sempre analisaram dados, mas antes, isso requeria muito mais tempo, era penoso e tinha alto custo, mas com o *Big Data*, tal processo ganhou enorme eficiência e rapidez.

A partir disto, aconteceram novas mudanças na produção jornalística, uma vez que ela está inserida em um sistema cada vez mais complexo. De acordo com a professora americana de ciência da computação, Melanie Mitchell (2009), o conceito de sistemas complexos se aplica aos processos que são auto organizados, emergentes e não triviais. Conectando essa concepção ao jornalismo, atualmente, ele pode ser entendido como um sistema formado por grandes redes de componentes sem controle central, com regras de operação

que crescem para o comportamento coletivo complexo, processam dados e informações sofisticadas e adaptam-se através do aprendizado e da evolução. Assim, apesar do jornalismo sempre ter sido baseado em dados, eles passaram a ter tanta complexidade em função do volume, da multiplicidade de fontes, da velocidade, entre outros fatores, que abriram espaço para as tecnologias estruturadas em algoritmos.

O algoritmo é um “guia” com equações matemáticas de comando, que define as etapas para a solução dos desejos operacionais inseridos nas máquinas a partir da realização de tarefas pré-definidas em suas linhas de código. Do ponto de vista computacional, ele opera em uma sequência de passos que pode ser denominada de seleção algorítmica (SAURWEIN; JUST; LATZER, 2015, p. 35). A partir de uma determinada seleção, o sistema transforma o dado que entra (*input*) em um resultado (*output*) para o usuário/jornalista, em uma velocidade e proporção que ele não é capaz de alcançar diretamente e sem a ajuda da tecnologia.

Ao pensar sobre o conceito de tecnologia, o filósofo Álvaro Vieira Pinto (2013, p. 166), que viveu antes da Internet das Coisas e do *Big Data*, lembra que o homem sempre utilizou técnicas para realizar uma finalidade a que ele se propõe, pois estas garantem maior rendimento através do emprego do menor esforço possível. Assim, novas técnicas são constantemente criadas com a ideia de satisfazerem as novas necessidades que são determinadas pelas sucessivas fases históricas. Seguindo esta concepção, alguns pesquisadores passaram a acreditar que as técnicas de exploração de dados digitais deveriam ganhar uma nova atuação, visando à geração de um valor adicional no século 21. Para o americano Kristian Hammond (2013), professor de Ciência da Computação e Jornalismo na *Northwestern University*, essa nova técnica seria o

desenvolvimento de um programa que não só fizesse a análise dos dados, mas que transmitisse rapidamente, os resultados encontrados em uma narrativa clara e concisa, em termos humanos ou em linguagem natural. Resumindo: o valor do *Big Data* não seria mais, portanto, apenas a exploração do dado propriamente dito, mas, sim, a possibilidade de construção de uma narrativa a partir dele. Com esta ideia e através do aprimoramento de um conjunto de tecnologias e técnicas, começou a ser possível produzir notícias automatizadas, com a necessidade da presença humana direta apenas na elaboração dos códigos de programação. Na automação de notícias, portanto, “o software importa os dados, usa algoritmos, extrai e organiza a informação, e depois a transforma em uma história com distribuição em larga escala” (NORMANDE, 2013, p. 370). Mesmo com textos mais burocráticos e sem criatividade, pelo menos até o momento, o uso da automação em várias redações americanas, europeias e chinesas, segue a lógica do que Robert Capps (2009), chama na revista *Wired*, de “revolução boa o suficiente” (*The good enough revolution*), onde tecnologias que promovem facilidades de uso, economias de tempo e redução de custos são adotadas mais facilmente, independentemente de reações iniciais adversas.

Com a automação, a tecnologia se espalhou por todos os segmentos sociais, indo muito além dos territórios comunicacionais profissionalizados. Mas, ao permitir a produção automática de notícias com *softwares*, conquistou posição ativa específica dentro dos sistemas de produção jornalística. Para compreender este fenômeno, optamos por usar a abordagem de Bruno Latour da Teoria Ator-Rede (TAR) por ter uma visão menos antropocêntrica do social. Segundo ela, um agregado não humano pode ter relação e conectar-se com outro humano, dentro de uma concepção do social como uma associação

momentânea e dinâmica. Assim, os algoritmos que redigem notícias são entendidos como atores - ou actantes - e como mediadores, pois eles alteram o ambiente que atuam. Para a Teoria Ator-Rede (TAR) ou Sociologia das Associações, qualquer coisa que modifica “uma situação fazendo diferença” (LATOURET, 2012, p. 108) é considerada um ator ou actante. Assim, para a TAR, “nenhuma ciência do social pode existir se a questão de o quê e quem participa da ação não for logo de início plenamente explorada” (LATOURET, 2012, p. 108). Quanto ao jornalismo, começar a investigação com os objetos “fornece uma nova janela para os contextos social, material e cultural que permeiam nosso mundo, cada vez mais tecnicamente obcecado” (ANDERSON; MAYER, 2015, p. 4, tradução nossa)⁴.

Inteligência Artificial e Geração de Linguagem Natural

John McCarthy lançou o termo inteligência artificial (IA) como tema da primeira reunião científica dedicada ao assunto, na Conferência no *Dartmouth College*, nos Estados Unidos, em 1956. A história da IA é cheia de altos e baixos. Na década de 50, os cientistas espalharam a ideia de máquinas capazes de realizar várias façanhas e de até superar os seres humanos em certos aspectos. Mas os primeiros robôs só executavam tarefas bem simples e, apesar dos avanços, a propaganda foi maior do que os feitos reais. Em 1980, houve novo entusiasmo quando o governo americano investiu em um robô que cruzasse a linha inimiga, fizesse o reconhecimento do terreno e voltasse para o quartel sozinho. Mas o robô acabou se perdendo. Apesar dos fracassos, ocorreram

⁴ No original: “provides a new window into the social, material, and cultural context that suffuses our increasingly technologically obsessed world” (ANDERSON; MAYER, 2015, p. 4).

alguns progressos nesta década, mas estes foram “sobrestimados, seguiu-se uma nova reação contrária e uma segunda hibernação, durante o qual o financiamento secou e se deu uma debandada desiludida” (KAKU, 2011, p. 99). Na década de 1990, nova euforia: o programa de computador *Deep Blue* da IBM derrotou o campeão mundial de xadrez, Garry Kasparov. Mas, e apesar de ser bom no jogo, o *Deep Blue* não foi bem em outros testes de inteligência (KAKU, 2011, p. 99). O entusiasmo retornou em 2011 quando milhões de espectadores acompanharam o poder da computação cognitiva com o *Watson*, desenvolvido também pela IBM e que venceu no *Jeopardy*, um famoso *game show* norte-americano. Dessa forma, na transição do final do século passado para o atual ocorreu uma consistente série de conquistas que permitiu o avanço dos sistemas de inteligência artificial, como os que possibilitam a produção de notícias automatizadas do momento.

De acordo com a definição contida na *Associated for Computing Machinery* (ACM)⁵, a Inteligência Artificial (IA) é um campo da Ciência da Computação e da Engenharia voltado para o comportamento inteligente e para a criação de artefatos que exibam este comportamento. A IA trabalha com sistemas baseados em conhecimentos, “sistemas especialistas, reconhecimento de padrões, aprendizado autônomo, compreensão de linguagem natural, robótica e outras” (KURZWEIL, 2007, p. 412). Para Nils Nilsson (2010, p.13, tradução nossa)⁶, autor de um extenso livro sobre a história das ideias e das conquistas da IA, a inteligência artificial pode ser resumida como “a qualidade que permite que uma entidade funcione adequadamente e com previsão no

⁵ A ACM é uma organização internacional, de caráter científico e educacional, que estabelece como um de seus objetivos o avanço das tecnologias de informação (<http://www.acm.org>).

⁶ No original: “*For me, artificial intelligence is that quality that enables an entity to function appropriately and with foresight in its environment*” (NILSSON, 2010, p. 13).

seu ambiente". Whitby (2004, p. 120), outro pesquisador norte-americano, adverte, no entanto, que é preciso tomar cuidado para não humanizar a ideia de inteligência, pois a IA é "diferente da inteligência natural. Entretanto, ainda assim é inteligência".

Interessado na geração de narrativas automatizadas, Kristian Hammond (2015, p. 12) explica que a inteligência, de forma geral, pode ser dividida em três fases (ver fig.1): separação de dados (compreensão), pensar sobre eles (raciocínio) e agir (ação). Segundo Hammond (2015, p. 13), a compreensão parte dos dados sobre o mundo, através de processamento de imagens e da filtragem e reconhecimento de palavras, etc. Já o raciocínio envolve: avaliação da situação; inferência (definição de que algo é verdadeiro porque, logicamente, deve ser); evidência (algo é verdadeiro baseado no peso da evidência disponível); planejamento para resolver uma tarefa; aprendizagem (construção de conhecimento a partir do que é examinado no conjunto de dados), etc. Depois que os dados foram compreendidos, processados e avaliados, chega o momento da ação.



Figura 1: Conjunto de competências dentro das etapas de IA

Fonte: Hammond (2015)

Muitos sistemas de inteligência artificial usam os três principais componentes do raciocínio (HAMMOND, 2015, p. 18): fazem avaliação, inferência - checam similaridades, categorizam, acumulam evidências - e predição. Com a explosão de dados (*Big Data* e IoT), os sistemas de inteligência artificial estão processando uma volumosa quantidade de exemplos que fazem com que eles possam agir com base em uma série expansiva de meios de aprendizagem. Hammond (2015, p. 30) conclui que não há magia neste processo, mas apenas objetivas aplicações de técnicas inteligentes nos dados. Na evolução tecnológica, estas técnicas e processos migraram para a geração de textos automatizados no jornalismo. Os algoritmos fazem parte de sistemas que trabalham com dados estruturados, os avaliam, fazem inferências

e agem através da produção das notícias. Para isso, eles trabalham com a Geração de Linguagem Natural (GLN), um campo de pesquisa que surgiu na década de 1950 e avançou muito com o aumento de dados digitais e com o aperfeiçoamento de técnicas de análise estatísticas.

A GLN é definida como um programa de computador que automaticamente produz exemplos da linguagem humana a partir de representações computacionais de informação (DÖRR, 2015, p. 4). O pesquisador da Universidade de Zurique, Konstantin Nicholas Dörr (2015, p. 5) aponta que a GLN é capaz de reduzir vários tipos de custos devido à automação e que, baseada na programação inicial do *software*, é capaz de tirar conclusões e identificar fatos interessantes, além de gerar texto de forma autônoma e muito mais rápida do que um jornalista é capaz.



Figura 2: Processo de geração de linguagem natural

Fonte: Reiter et al. (2005) e Dörr (2015)

O processo de geração pode ser dividido em três estágios (fig.2). A primeira etapa planeja o conteúdo e a estrutura, assim como decide qual informação vai mostrar no texto a partir dos dados numéricos (REITER et al., 2005, p.142). Pegando a geração de um texto sobre previsão do tempo, por exemplo, observamos que na fase do planejamento pode haver a decisão de optar por mostrar mudanças na direção do vento, descartando os dados sobre

a velocidade do mesmo. Na segunda etapa da GLN, “os algoritmos aplicam operações estatísticas que selecionam elementos de um conjunto de dados básicos e atribuem relevância a eles”. O sistema também “deve decidir quais estruturas linguísticas (palavras, sintaxe, frases) devem ser usadas para comunicar a informação desejada” (DÖRR, 2015, p. 6, tradução nossa).⁷ Além disso, na fase do micro planejamento é tomada a decisão de como as informações nas sentenças serão distribuídas, se elas serão curtas ou longas, etc. (REITER et al., 2005, p.142). Na etapa de realização da redação, o texto é gerado de acordo com as definições das outras duas fases. Resumindo esses momentos, durante o processamento de geração de linguagem natural é preciso definir o que dizer, quando e como fazê-lo, realizando estrutura similar que se prepara para o *lead* jornalístico. Os pesquisadores Márcio Carneiro dos Santos (2016b) e Tacyana Arce (2009) lembram que a possibilidade de automação da notícia já havia sido apontada anteriormente aqui no Brasil, em 1997, por Nilson Lage. Arce, citando Lage, diz que o *lead* pode ser analisado como função, no sentido matemático. Desta forma, ele poderia ser resumido em uma sentença geral: $F(x, y, z, \dots)$. Onde F representa o verbo e x, y, z, \dots , os argumentos, “isto é, o sujeito x (argumento externo) e complementos y, z, \dots (argumentos internos do verbo)” (ARCE, 2009, p. 9). Dörr (2015, p. 13, tradução nossa) explica que as empresas começam, em geral, a automatização de textos mais nas áreas de esportes e finanças porque são assuntos onde os “dados estruturados estão disponíveis em uma quantidade escalável. Além disso, o idioma, regras e configurações podem ser facilmente definidas em termos de

⁷ No original: “algorithms apply statistical operations to select elements from a basic data set and assign relevância to them. The NGL system must decid wich linguistic structures (words, syntax, sentences) should be used to communication the desired information” (DÖRR, 2015, p. 6).

programação”⁸. Na figura 3 tem um exemplo de notícia automatizada.

EPS Estimates Down for J.M. Smucker In Past Month



Narrative Science,

FULL BIO ▾

Opinions expressed by Forbes Contributors are their own.

Over the past three months, the consensus estimate has sagged from \$1.25. For the fiscal year, analysts are expecting earnings of \$5.75 per share. A year after being \$1.37 billion, analysts expect revenue to fall 1% year-over-year to \$1.35 billion for the quarter. For the year, revenue is expected to come in at \$5.93 billion.

A year-over-year drop in revenue in the fourth quarter broke a three-quarter streak of revenue increases.

The company has been profitable for the last eight quarters, and for the last four, profit has risen year-over-year by an average of 16%. The biggest boost for the company came in the third quarter, when profit jumped by 32%.

Earnings estimates provided by Zacks.

Narrative Science, through its proprietary artificial intelligence platform, transforms data into stories and insights.

Figura 3: Autoria da *Narrative Science* no *site da Forbes.com*

Fonte: *Forbes.com*⁹

Após esta breve exposição, cabe apresentar uma observação quanto à geração de conteúdo. Segundo a análise do professor da Faculdade de Engenharia Elétrica e de Computação da Unicamp, Ricardo Gudwin (2017), uma

⁸ No original: “structured data are available in a scalable amount. Additionally, the language, rules, and settings can be easily defined in terms of programing” (DÖRR, 2015, p.13).

⁹ Disponível em: <<https://www.forbes.com/sites/narrativescience/2015/10/12/eps-estimates-down-for-j-m-smucker-in-past-month/#2c07c5e27595>>. Acesso: 07 fev. 2017.

parte dos sistemas que geram texto automaticamente não possui de fato um “entendimento” da linguagem que é gerada. Nestes casos, eles são, simplesmente, processadores de textos que operam ferramentas da linguística computacional que geram um texto semelhante à linguagem natural e que podem até iludir quem interage com eles, mas isto não quer dizer que estão de fato compreendendo o que geram.

Neste artigo não focamos nas divisões de tarefas entre jornalistas e algoritmos de inteligência artificial responsáveis pela automação. Mas apontamos o surgimento de uma nova tarefa para alguns profissionais nas redações: o “treinamento” dos algoritmos (LINDEN, 2016, p.12). Assim, caberá a uma parte dos jornalistas conhecer e dominar a forma de agir dos programas de automatização, o que eles devem buscar, que palavras usar e quais estarão proibidas, determinar as possibilidades de sinônimos e como estruturar as notícias que o *software* vai produzir, de acordo com estilo da empresa de jornalismo envolvida. O pesquisador finlandês Carl-Gustav Linden (2016, p. 12) vê nessa tarefa algo nem sempre agradável para o profissional e levanta a possibilidade de que alguns jornalistas sejam reduzidos a simples operadores de máquinas (2016, p. 13).

Notícias Automatizadas

O emprego de algoritmos para produzir notícias automaticamente já está ocorrendo em várias redações, seja através de plataformas desenvolvidas por empresas de tecnologia ou mesmo internamente por equipes dos próprios grupos de comunicação. Ainda que o momento atual represente uma primeira onda de automação no jornalismo (LINDEN, 2016), identificamos, durante a revisão bibliográfica, que notícias estão sendo produzidas automaticamente em

pelo menos dez países: Estados Unidos, Alemanha, Reino Unido, França, Suécia, Dinamarca, Noruega, Rússia, China e Coreia do Sul. Entre os grupos que oferecem notícias automatizadas, podem ser citados, por exemplo, o alemão SID (*Sports Information Service*); os americanos *Yahoo*, *Los Angeles Times*, *Washington Post*, *Bloomberg*, *Forbes*, *ProPublica*, *Associated Press*; o francês *Le Monde*; o sueco *Mittmedia*; o grupo *MeteoGroup* do Reino Unido, as agências chinesa *Xinhua* e a russa *Interfax*, entre outras.

Além da automação só ser usada exclusivamente a partir de dados estruturados, ela apenas produz, até agora, textos de notícias, mas não os de reportagens. Para compreender esta distinção, entende-se notícia como a “produção da informação primária sobre evento concreto e objetivo. Já a reportagem é resultado de operação analítica ou crítica da realidade, o que exige alto grau de subjetividade, algo, portanto, pouco propício à automação” (ARCE, 2009, p. 4). Cada etapa da produção pode ser automatizada agora: da apuração e redação de notícias à distribuição e circulação, sem contar que os algoritmos podem fazer dezenas, centenas e até milhares de artigos rapidamente (DALEN, 2012, p. 649), além de permitirem colocar automaticamente uma *tag*¹⁰ no conteúdo para que ele possa ser acessado por consumidores de notícias com determinado perfil. Tal possibilidade permite gerar aumento da receita através da personalização, o que pode ter forte impacto no modelo de “negócios” do jornalismo (fig.4).

¹⁰ *Tag*, em inglês, significa etiqueta. Na internet, ela funciona como um marcador, que facilita a localização da informação.

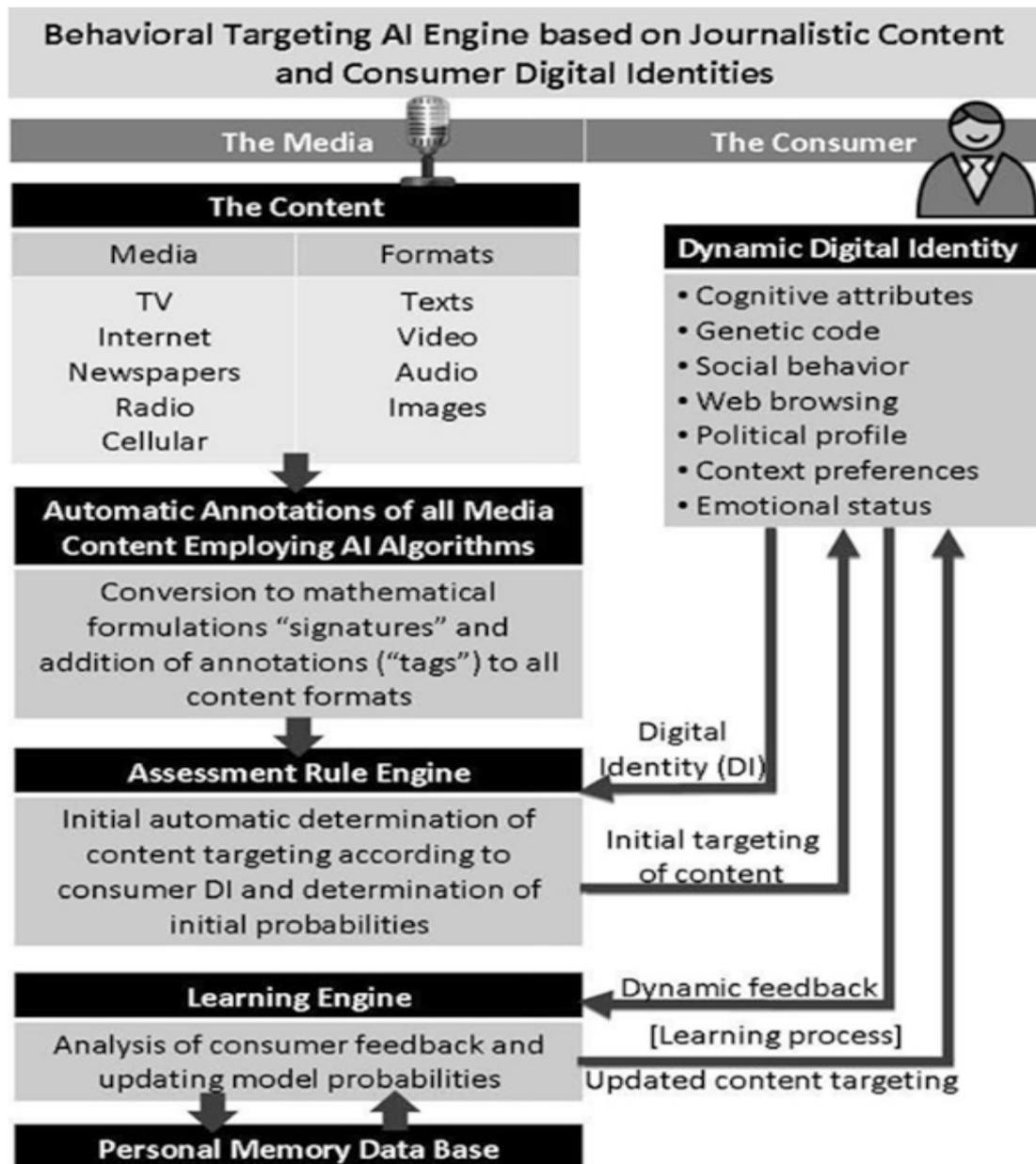


Figura 4: Esquema de segmentação de conteúdo jornalístico
Fonte: Latar (2015)

A possibilidade de personalização altera drasticamente as condições de produção e recepção das matérias jornalísticas (CARLSON, 2015, p.425). E pode

ter inúmeras consequências sociais, como a ampliação do que Pariser (2012) chama de “bolha de filtros”. “Quando a tecnologia passa a nos mostrar o mundo, acaba por se colocar entre nós e a realidade, como a lente de uma câmera (PARISER, 2012, p. 18). O autor completa,

os filtros de personalização servem como uma espécie de autopropaganda invisível, doutrinando-nos com as nossas próprias ideias, amplificando nosso desejo por coisas conhecidas e nos deixando alheios aos perigos ocultos no obscuro território do desconhecido. Na bolha de filtros, há menos espaço para os encontros fortuitos que nos trazem novas percepções e aprendizados (PARISER, 2012, p. 19).

Conclui-se, de forma resumida, que a automação está sendo vista, nesta primeira fase, como:

- a) Útil para histórias de rotina e tópicos repetitivos (GRAEFE, 2016). No entanto, o texto é mais burocrático, não é criativo e pode ter problemas com metáforas e contextos culturais diversos. Os algoritmos não fazem pergunta “fora da caixa” (LATAR, 2015), reduzem a diversidade, complexidade, curiosidade e imaginação (CARR, 2015) e eliminam empregos de rotina do jornalista (LINDEN, 2016);
- b) Válida quando a velocidade é essencial (GRAEFE, 2016), além de satisfazer a necessidade de imediatismo (LINDEN, 2016). Ela economiza tempo de produção de notícias e também, ao mesmo tempo, pode postar matérias rapidamente em redes sociais (LATAR, 2015). No Le Monde, por exemplo, os algoritmos redigiram 150 mil textos curtos sobre as eleições departamentais de 2015 num período de 4 horas: ou

- seja, 625 notícias por minuto e 10 para cada segundo¹¹;
- c) Permite expansão da cobertura e aumento da receita através da personalização, (LECOMPTE, 2015; LATAR, 2015). Por outro lado, aumenta a pressão para adaptar conteúdo de acordo com retorno econômico (LATAR, 2015);
 - d) Identifica fatos despercebidos em enormes volumes de dados e gera decisões e processos editoriais mais eficientes ao fazer previsões e detectar tendências (LINDEN, 2016; LATAR, 2015);
 - e) Libera repórteres para tarefas que exigem mais qualificação, análise, investigação e contextualização (DALEN, 2012; LECOMPTE, 2015). Como os algoritmos podem realizar funções de rotina típicas de jornalistas em começo de carreira, há um risco de diminuição deste tipo de trabalho, sem contar que, segundo Linden (2016, p.13), isso pode implicar em novas formas de aquisição de conhecimento tático por jovens profissionais, o que deve impactar na forma de ensino e formação.

Ainda que não tenham relatos sobre demissões nas redações ocorridas em função da automação, a questão levanta temores, principalmente, se levarmos em consideração um estudo feito por pesquisadores da *Oxford University* e da *Delloite* sobre 702 profissões. Segundo eles, cerca de 47% dos postos de trabalho dos Estados Unidos, nas próximas duas décadas, estarão ameaçados pelo o que os autores chamam de *computerisation*, isto é, pela informatização. (FREY; OSBORNE, 2013). Eles criaram uma metodologia para identificar o percentual de suscetibilidade de perda de emprego. Repórteres e

¹¹ Disponível em: <<http://mediashift.org/2016/07/upsides-downsides-automated-robot-journalism/>>. Acesso: 28 ago.2016.

correspondentes ocupam a 177^a. posição na lista dos 702 estudados, sendo que quanto mais perto da 1^a. posição, menos chance existe de perder o trabalho em função da tecnologia. Jornalistas têm, evidentemente, várias vantagens em relação à automação. Mas segundo o israelense Noam Latar (2015, p. 67), eles precisam se adaptar porque vão competir cada vez mais com os algoritmos na produção de notícias.

Portanto, ao mesmo tempo que sinaliza alguns benefícios, a automação no jornalismo demonstra desvantagens e motivos de preocupação. Matt Carlson (2015) vê na automatização o que Capps (2009) chama de *Good Enough Revolution* e que já foi citada na introdução deste artigo. Ser bom o suficiente representa um novo sistema de valores. A automatização pressupõe a ascensão de uma tecnologia que não visa garantir a alta qualidade de análise e de texto, muito menos de investigação no jornalismo. Mas ela é suficientemente boa para ser usada porque propicia uma funcionalidade adequada e mais barata (CARLSON, 2014, p. 425) para as empresas de comunicação, além de atender a demanda de imediatismo reforçada com a internet e com os dispositivos móveis (LINDEN, 2016).

Os desafios em torno das notícias automatizadas são grandes e é de suma importância discutir as prováveis consequências de sua utilização para evitar cair em armadilhas, resultantes de uma atitude passiva em relação a elas. Como lembram Erik Brynjolfsson e Andrew McAfee (2015, p. 12), os problemas “não são intransponíveis” quando existe diálogo, mas as tecnologias “não vão reparar a si mesmas” sem a devida intervenção humana.

Considerações Finais

A tecnologia sempre foi usada pelo jornalismo como instrumento, mas a partir da automação de notícias, ela também passou a produzi-lo, ainda que dentro de certas limitações como a possibilidade de ser empregada apenas quando existem dados estruturados. A automatização está sendo implementada para atender determinadas necessidades liberadas com a digitalização, o que impôs numerosas alterações no ecossistema do jornalismo – desde a demanda por mais notícias e fluxo contínuo de produção aos problemas gerados pela crise no modelo de “negócios”. Algoritmos extremamente sofisticados conseguem processar dados em alta velocidade como nenhum ser humano é capaz. Eles também transformam esses dados em narrativa com uma rapidez inigualável. Essas vantagens estão sendo vistas como suficientemente boas e interessantes para o modelo de “negócios” do jornalismo. Mas a automação também gera problemas que precisam de uma maior análise, desde a questão da bolha de filtros que envolve o consumidor em um mundo feito sob medida sem a necessária dose de diversidade até o papel dos jornalistas.

Pelo tema ser recente e rico, não pretendemos esgotar o universo de questões que envolvem a automação. Ela tem várias implicações e desdobramentos éticos que não podem ser esquecidos, mas que não fizeram parte do objeto focal deste artigo. Outrossim, nossa ideia, sem nenhuma pretensão determinista, foi a de apontar que a automatização já é realidade em várias redações e que ela está redefinindo o fazer jornalístico. Durante a revisão bibliográfica observamos que notícias estão sendo redigidas automaticamente em pelo menos dez países. O Brasil, por enquanto, não é um deles. Mas, se forem analisados o desenvolvimento de outras tecnologias nos continentes

norte-americano e europeu, é possível imaginar que existe uma chance real da automatização de notícias se tornar uma prática no Brasil. Apesar de uma série de mudanças no jornalismo, a missão principal de uma atividade jornalística isenta e plural não se alterou e continua sendo um serviço essencial para a democracia. Acreditamos que a tecnologia pode adicionar uma parte importante neste intento.

Referências

ANDERSON, C. W; MAYER, Juliette. **Objects of journalism and the news. Journalism**, 16 v., 2015.

ARCE, Tacyana. O lead automatizado: uma possibilidade de tratamento da informação para o jornalismo impresso diário. **Revista Exacta**, Belo Horizonte, v.2, n. 3, 2009.

BRYNJOLFSSON, Erik; MCAFEE, Andrew. **A segunda era das máquinas: trabalho, progresso e prosperidade em uma época de tecnologias brilhantes**. Rio de Janeiro: Alta Books, 2015.

BUNGE, Mario. **Dicionário de filosofia**. Tradução de Gita K. Guinsburg. 1. reimp. São Paulo: Perspectiva, 2012.

CARLSON, Matt. The robotic reporter: automated journalism and the redefinition of labor, compositional forms and journalistic authority. In: LEWIS, Seth C. (Org.). **Digital Journalism**. v.3, n.3. New York: Taylor&Francis Online, 2016.

CAPPS, Robert. The good enough revolution: when cheap and simple is just fine. **Wired**, 24 ago. 2009. Disponível: <<https://www.wired.com/2009/08/ff-goodenough/>> Acesso: 10 ago. 2016.

DALEN, Arjen van. The Algorithms Behind the Headlines: How machine-written news redefines the core skills of human journalists. **Journalism Practice**. Volume 6, Issue 5-6. New York: Routledge, 2012.

DAVENPORT, Thomas H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo, Futura, 1998. Disponível em: <<http://amormino.com.br/livros/20141114-ecologia-informacao.pdf>>. Acesso em: 10 de ago. 2015.

DIAKOPOULOS, Nicholas. Algorithmic accountability reporting: on the investigation of black boxes. **Town Center for Digital Journalism**. 2013. Disponível em: <http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf>. Acesso: 10 abr. 2016.

DÖRR, Konstantin Nicholas. Mapping the field of algorithmic journalism. 2015. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/21670811.2015.1096748>>. Acesso: 24 set.2016.

FLORIDI, Luciano. The language of information. In: _____ **Information**: a very short introduction. New York: Oxford University Press, 2000. p. 19-36.

GRAEFE, Andreas. Guide to automated journalism. **Town Center for Digital Journalism**. Jan. 2016. Disponível em: <<http://towcenter.org/research/guide-to-automated-journalism/>> Acesso em 08 abr. 2016.

GUDWIN, Ricardo. **Geração de linguagem natural**. [mensagem pessoal]. Mensagem recebida por <krishmacarreira@gmail.com> em 31 jan.2017.

KAKU, Michio. **A física do futuro**: como a ciência moldará o mundo nos próximos cem anos. Tradução de Maria Carvalho e João C. S. Duarte. Lisboa: Editorial Bizâncio, 2011.

KURZWEIL. **A era das máquinas espirituais**. Tradução de Fábio Fernandes. 2.reimp. São Paulo: Aleph, 2007.

LATAR, Noam. The Robot Journalism in the Age of Social Physics: the end of human journalism? **The New World of Transitioned Media**. *Springer*, 2015.

LATOUR. Bruno. **Reagregando o social**: uma introdução à teoria Ator-Rede. Tradução de Gilson César Cardoso de Sousa. Salvador: Edufba, 2012.

LECOMPTE, Celeste. Automation in the Newsroom. **Nieman Foundation**, 1º. set. 2015. Disponível em: < <http://niemanreports.org/articles/automation-in-the-newsroom> >. Acesso em: 02 mar. 2016.

LINDEN, Car-Gustav. **Decades of automation in the newsroom**: why are there still so many jobs in journalism? Mar. 2016.

HAMMOND, Kristian. The value of big data isn't the data. *Harvard Business Review*, mai. 2013. Disponível em: < <https://hbr.org/2013/05/the-value-of-big-data-isnt-the>>. Acesso: 22 fev.2016.

_____. **Practical artificial intelligence for dummies**. 2015. Disponível em: <http://gunkelweb.com/coms493/texts/AI_Dummies.pdf>. Acesso: 28 mai. 2016.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth (2013). **Big Data**: a revolution that will transform how we live, work and think. London: John Murray, 2013.

MITCHELL, Melanie. **Complexity**: a guide tour. New York: Oxford University Press, 2009.

NILSSON, Nils J. **The quest of artificial intelligence**: a history of ideas and achievements. Standard University. 2010.

NORMANDE, Naara. A automatização da narrativa jornalística. **Estudos em Comunicação**, n.13, 2013, p.363-378. Disponível em: <<http://www.ec.ubi.pt/ec/13/pdf/EC13-2013Junho-14.pdf>>. Acesso: 10 abr. 2016.

OSBORNE, Michael A.; FREY, Carl Benedikt. **The future of employment**: how susceptible are jobs to computerisation? Oxford University, sep. 17, 2013.

PARASIE, Sylvain; DAGIRAL, Eric. Data-driven journalism and the public good: "computer-assisted-reporters" and "programmer-journalists" in Chicago. **New Media & Society**. 2012. Disponível em: <<http://www.blogg.sh.se/datajournalistik/wp-content/uploads/2013/02/New-Media-Society-2012-Parasie-1461444812463345.pdf>>. Acesso: 17 set. 2016.



ISSN nº 2447-4266

Vol. 3, n. 3, Maio. 2017

DOI: <http://dx.doi.org/10.20873/uft.2447-4266.2017v3n3p60>

PARISER, Eli. **O filtro invisível**: o que a internet está escondendo de você. Tradução de Diego Alfaro. Rio de Janeiro: Zahar, 2012.

PINTO, Álvaro Vieira. **O conceito de tecnologia**. 1v, 2 reimp. Rio de Janeiro: Contraponto, 2013.

PRIMO, Alex; ZAGO, Gabriela. Who and what do journalism? An actor-network perspective. **Digital Journalism**, 2014. Vol. 3, no. 1, p. 38-52.

SANTOS, Márcio Carneiro dos. _____ . **Comunicação digital e jornalismo de inserção**: como big data, inteligência artificial, realidade aumentada e internet das coisas estão mudando a produção de conteúdo informativo. São Luis: Labcom Digital, 2016a.

_____ Automated narratives and journalistic text generation: the lead organization structure translated into code. **Brazilian Journalism Research**: Revista da Associação Brasileira de Pesquisadores em Jornalismo (SBPJor), Brasília, v.12, n.1, 2016b. P. 150-175. Disponível em: < <https://bjr.sbpjor.org.br/bjr/article/view/921>>. Acesso: 30 jun.2016.

SAURWEIN, Florian; JUST, Natascha; LATZER, Michael. 2015. **Governance of algorithms**: options and limitations. Set. 2015. Disponível em:< <http://www.mediachange.ch/media/pdf/publications/GovernanceOfAlgorithms.pdf>>. Acesso: 2 ago. 2016.

WHITBY, Blay. **Inteligência Artificial**: um guia para iniciantes. Tradução: Claudio Blanc. São Paulo: Madras, 2004.