

DOSSIÊ

A OPERAÇÃO JORNALÍSTICA GUIADA POR DADOS:

reconfigurando a percepção de valor
da notícia com algoritmos, inteligência
artificial e personalização estendida



MÁRCIO CARNEIRO DOS SANTOS

Universidade Federal do Maranhão, São Luís – Maranhão – Brasil

ORCID: 0000-0002-0663-1300

DOI: 10.25200/BJR.v16n3.2020.1295

Recebido em: 03/04/2020

Desk Review em: 04/05/2020

Editores do Desk Review: Laura Storch e David Domingo

Revisado em: 17/08/2020

Aceito em: 03/11/2020

RESUMO – Discute-se a reconfiguração estratégica da utilização de dados dentro das organizações jornalísticas, através de algoritmos e soluções de inteligência artificial, entre elas especificamente o aprendizado de máquina, como alternativa para elevação da percepção de valor sobre o produto informativo, reduzida pela fragmentação das audiências, explosão de emissores, concorrência indireta das grandes empresas de tecnologia e transformações do ecossistema de meios digitais. Tal abordagem propõe a identificação de padrões de interesse, a predição de engajamento social e a alocação de recursos para coberturas como formas de expandir o atual nível de personalização oferecido aos consumidores de notícia.

Palavras-chave: Algoritmos. Inteligência artificial. Aprendizado de máquina.

**THE DATA-GUIDED JOURNALISTIC OPERATION:
reconfiguring the perception of news value with algorithms, artificial
intelligence and extended personalization**

ABSTRACT – This article discusses the strategic reconfiguration of data in news organizations

through the use of algorithms and artificial intelligence, more specifically machine learning, to increase the idea of value around information, which has suffered from fragmented audiences, a massive increase in the number of broadcasters, indirect competition from large technology companies, and changes to the digital media ecosystem. We propose to identify patterns of interest, predict social engagement, and allocate resources for new coverage as forms for increasing the current level of personalization offered to news consumers.

Key words: algorithms, artificial intelligence, machine learning

LA OPERACIÓN PERIODÍSTICA GUIADA POR DATOS: reconfigurando una percepción del valor de la notificación con algoritmos, inteligencia artificial y personalización estendida

RESUMEN – Se discute la reconfiguración estratégica del uso de datos en las organizaciones periodísticas a través de algoritmos y soluciones de inteligencia artificial, específicamente, el aprendizaje automático como una alternativa para aumentar la percepción de valor sobre el producto de información, el cual se reduce por la fragmentación de las audiencias, el aumento en la cantidad de emisores, la competencia indirecta de grandes compañías tecnológicas y las transformaciones del ecosistema de medios digitales. Tal enfoque propone la identificación de patrones de interés, la predicción del compromiso social y la asignación de recursos para la cobertura como formas de expandir el nivel actual de personalización ofrecido a los consumidores de noticias.

Palabra clave: Algoritmos. Inteligencia artificial. Aprendizaje automático

1 Introdução

Nas últimas décadas, muito material acadêmico foi produzido sobre um conjunto de conceitos relacionados à utilização de recursos computacionais, em diferentes níveis, dentro do processo de produção de conteúdo jornalístico. Reportagem assistida por computador (RAC), jornalismo digital em bases de dados (JDBD), jornalismo guiado por dados (JGD ou *data driven journalism*), jornalismo computacional e jornalismo robótico são alguns deles.

Os três primeiros estão ligados ao contínuo crescimento do volume de informações a que estão expostos os jornalistas, chegando à condição de excesso de dados disponíveis no ambiente digital e o impacto desse fato na construção da notícia.

A ideia mais utilitária do computador, como ferramenta a ser operada pelo jornalista, para o apoio às diversas etapas do seu trabalho, está na base da RAC.

Em outra vertente, toda uma literatura ligada à utilização de bases de dados como suporte do trabalho das redações no seu *backend*, ou seja, na operação interna, que roda por trás do que é distribuído, nos sistemas de gerenciamento de conteúdo, ferramentas de publicação e controle de redações, foi consolidada em termos teóricos através do paradigma JDBC (Barbosa, 2007, 2008, 2009, 2011; Fidalgo, 2004, 2007; Machado, 2006) citando as bases de dados “como definidoras da estrutura e da organização, bem como da composição e da apresentação de conteúdos de natureza jornalística” (Barbosa & Torres, 2013, p.154).

O conceito de dados abertos¹ (*open data*), a transposição e consolidação de enormes bases de dados públicas para portais de transparência e similares, além das alterações no conjunto de leis nacionais ligadas ao acesso à informação e ao controle sobre os gastos públicos, reforçaram a ideia de que os repositórios digitais fossem vistos, pelos jornalistas, como fontes a serem também entrevistadas e lugares para identificar novas histórias, que esses dados poderiam ajudar a contar. Muito do que se chama hoje de jornalismo guiado por dados parte dessas premissas.

Mais próximos do conceito de jornalismo computacional, trabalhos sobre a automatização do processo de produção, a ponto de termos *softwares* capazes de redigir textos jornalísticos, são encontrados em Carreira (2017), que fala de um jornalismo feito por não humanos, e Višnovský et al. (2019), que propõem um robojornalismo, resultado da interconexão entre informática, estatística e reportagem. Outros textos usam termos como jornalismo automatizado e jornalismo robótico (Maier, 2002; Dalen, 2012; Carlson, 2014; Clerwall, 2014; Lewis & Usher, 2014; Latar, 2015; Santos, 2014; Santos, 2016a; Graefe, 2016) ao tratarem de processos semelhantes.

A percepção interna sobre as mudanças no processo de produção a partir do uso de algoritmos e automatização é outra vertente sobre o tema, encontrada em trabalhos como os de Rubio e Ruiz (2020) e López et al. (2018).

Coddington (2014) ressalta algumas diferenças entre esses conceitos:

- a) A reportagem assistida por computador (RAC), conceito

mais antigo, baseado nas ideias de Meyer (1973) de um jornalismo de precisão, fundamentado na proposta de uma atividade desenvolvida a partir de técnicas da pesquisa acadêmica.

b) O jornalismo guiado por dados (JGD), onde a interação com os recursos computacionais e a abordagem quantitativa são o eixo central na proposição das pautas e no trabalho dos jornalistas que são treinados para trabalhar com planilhas, bases de dados, visualizações e programação.

c) O jornalismo computacional (JC) onde a aproximação com o desenvolvimento de algoritmos, a automatização de etapas do processo de produção e a eventual substituição do jornalista pela máquina, em alguns momentos, são características importantes.

Diakoupolous (2011) diz que:

Jornalismo computacional é a aplicação de computação e pensamento computacional às atividades de jornalismo, incluindo coleta de informações, organização e produção de sentido, comunicação e apresentação, disseminação e resposta pública a informações de notícias. (Diakoupolous, 2011, p.1, tradução do autor²).

O foco de todas essas abordagens, entretanto, não contempla outro tipo de utilização de dados e algoritmos que se conecta, potencialmente, com a operação da organização jornalística como um todo.

Essa angulação, que é a intenção deste texto, vai além dos aspectos ligados à produção das notícias, sendo mais orientada ao potencial de utilização de ferramentas computacionais nas relações que a organização jornalística estabelece com o ambiente externo que a envolve, incluindo usuários, leitores, consumidores, assinantes, anunciantes, agências e mercado em geral, abarcando tópicos como modelos de negócio e percepção de valor sobre o produto que entrega, delineando uma abordagem não diretamente ligada ao conteúdo jornalístico, e sim, à economia da comunicação.

Essa perspectiva enquadra-se em uma área de interseção entre comunicação em geral, jornalismo, marketing, ciência de dados e tecnologias emergentes, caracterizando-se como eminentemente interdisciplinar e complexa.

Nosso objetivo é analisar, mesmo que de forma exploratória, a transição de uma utilização de dados e algoritmos que consideramos incipiente, no caso das organizações jornalísticas, para um patamar alcançado (ou pelo menos mais intensamente buscado) por outros

tipos de indústria, onde dados e algoritmos têm se tornado fator estratégico de fundamental centralidade, não só para conquista ou manutenção de mercados como para sua própria sobrevivência.

Neste texto, discutimos, utilizando uma proposição teórica geral e três possíveis repercussões aplicadas, apenas a título de exemplificação deste caminho, a reconfiguração de um tipo de insumo, sempre abundante nas redações, mas historicamente subutilizado: os dados. Isso mediante a implementação de uma cultura interna orientada à geração de valor a partir da informação, não apenas na produção da notícia (o que sempre existiu), mas em diversos aspectos do funcionamento da indústria do jornalismo, incluindo suas rotinas operacionais, sua relação com os consumidores, o design de seus produtos e suas formas de distribuição.

2 A indústria da notícia e sua crise

No mundo analógico, o modelo de negócio do jornalismo impresso tradicional baseava-se em três fontes: publicidade, venda em banca e comercialização de assinaturas. O jornalismo eletrônico do rádio e da TV, pela impossibilidade material das duas últimas formas, cresceu a partir da entrega de um conteúdo aparentemente gratuito, entremeado com *breaks*, intervalos comerciais, onde o anunciante, de forma indireta, pagava (e paga até hoje) para falar com o público, entre os blocos das notícias.

Sobre esse terreno estável, décadas de atividade jornalística foram construídas basicamente porque a geração de receita do negócio funcionava a contento, produzindo recurso que pagava a operação do processo de produção da informação e dividendos financeiros e políticos, através do exercício de uma influência considerável, a partir da entrega de um conteúdo massivo e com pouca concorrência.

Os grandes grupos da comunicação brasileira, de um jeito ou de outro, cresceram a partir dessa situação de centralidade, onde o controle dos canais de difusão de informação significava lucro e poder.

Com a chegada da internet, a indústria da notícia imaginou que tudo seria como antes, ou talvez até melhor. O novo meio seria apenas mais um canal de entrega onde o modelo consagrado continuaria operando com a mesma eficiência, atingindo novos públicos e gerando mais receita, ajudando a expandir a força das organizações que controlavam os canais tradicionais.

O tempo mostrou que não foi bem assim que aconteceu. A transição digital alterou o paradigma de consumo da informação, baseado em escassez, centralidade, homogeneidade e unidirecionalidade, para outro bem diferente, onde a abundância de conteúdo, a descentralização das fontes com a explosão de emissores, a fragmentação das audiências e as possibilidades de interação e interferência no fluxo, antes apenas de um para muitos, são as características identificáveis, no atual estágio de um processo de transformação acelerado e ainda incompleto.

Os grandes veículos, mesmo antes da internet, sempre foram enormes *hubs*, concentradores de informação e atenção, inicialmente controlando fluxos unidirecionais de informação, no sentido do emissor para seu público e, no sentido inverso, de total atenção da audiência para com o conteúdo recebido.

No ambiente digital, a estrutura das redes que conectam pessoas e veículos de mídia foi alterada, os obrigando a lidar com a interação bidirecional, onde informação e atenção assumiram um delicado equilíbrio, em um sistema complexo no qual a hegemonia da grande mídia teve sua solidez reduzida. Tanto a informação quanto a atenção passaram a trafegar nos dois sentidos de cada vínculo ligando pessoas e meios.

A informação começou a fluir de forma mais intensa dos usuários para os veículos através do recente poder dos primeiros em se tornarem, eles mesmos, emissores, capazes de produzir conteúdo em texto e audiovisual, utilizando na maioria das vezes apenas um celular e se tornando, em diversos arranjos, não apenas fontes, mas colaboradores, avaliadores, parceiros na difusão e compartilhamento do conteúdo e, indiretamente, geradores de métricas e *insights* sobre preferências hoje indispensáveis em qualquer redação.

Esse contexto fez com que a atenção dos meios se direcionasse de forma mais intensa para os consumidores de informação. Não que isso não ocorresse antes, entretanto, além de simplesmente garantir que o conteúdo chegue à audiência, agora, é necessário compreendê-la em profundidade e em tempo real, monitorando todo o engajamento possível, não só pelos canais tradicionais de contato, mas pelas redes sociais, plataformas de vídeo e qualquer outra porta digital onde seja viável extrair dados sobre quem está do outro lado.

O impacto dessas mudanças na indústria da notícia foi devastador e a capacidade das organizações em se adaptar ao novo cenário continua sendo testada diariamente.

Redações foram fechadas ou drasticamente reduzidas. Retomou-se então a discussão sobre uma crise do jornalismo identificada em diversas frentes: ética, técnica e talvez, ao nosso modo de ver, a mais letal, a crise de um modelo de negócio que, mesmo tentando se readaptar, foi se mostrando ineficiente em virtude das transformações significativas no ambiente onde operam as organizações jornalísticas.

A aplicação de tecnologias emergentes como inteligência artificial – IA (Coppin, 2010), internet das coisas – IoT (Santos, 2016b; Santos, 2016c), realidade virtual (Santos, 2019a), realidade aumentada (Santos, 2015) e, principalmente, *big data* (Santos, 2019b), ou seja, um conjunto de ferramentas e processos para lidar com dados caracterizados por seu volume, variedade e velocidade de produção, todas implementadas via código sob o controle de algoritmos, é incipiente dentro da operação das organizações jornalísticas, em estágio bem diferente de outras indústrias.

Defendemos aqui a hipótese de que o cerne da crise da indústria do jornalismo está na redução drástica da percepção de valor, por parte dos consumidores, sobre o produto jornalístico. O ambiente anterior de escassez que nos fazia sair de casa para ir a uma banca comprar o jornal do dia ou nos paralisava diante da TV, sempre nos mesmos horários, para garantir nosso quinhão de informação, de acordo com a grade de programação da emissora, foi se esvaziando ao longo do tempo. A abundância de fontes e conteúdo informativo amplamente disponível via internet, redes sociais e aplicativos diversos, todos facilmente acessáveis pelos celulares, cada vez mais populares, fez com que a necessidade percebida do produto jornalístico tradicional fosse se perdendo.

A reação natural da indústria foi enfrentar a fragmentação das audiências e a perda de receitas e centralidade a partir de um conjunto de tentativas de maior ou menor sucesso dependendo da região, tamanho do negócio, nível de credibilidade e capacidade de rever modelos e processos de operação.

A conversão da base de assinantes do produto físico para o digital, os *paywalls*, as conexões com as plataformas de mídias sociais, a reconfiguração dos sites e tentativas mais inovadoras como a produção de conteúdo imersivo são apenas algumas das diversas ações desenvolvidas pelas organizações jornalísticas que buscam se realinhar diante de um cenário hostil.

A busca fundamental pela informação passou a ser atendida por diversos outros mecanismos, mesmo que com isso a desinformação, as *fake news*, e todo um conjunto de novos problemas fossem sendo introduzidos no ecossistema dos meios.

Apesar da impossibilidade, pelas limitações deste texto, de discutir o tema da produção deliberada de desinformação, há que se ressaltar o fato de que as transformações econômicas, sociais e tecnológicas registradas no ambiente digital o tornaram alvo da ação de agendas diversas, orientadas a influenciar, tutelar ou conduzir o debate público, utilizando, entre outros recursos, o uso de ferramentas de automação e uma compreensão da lógica interna de funcionamento dos algoritmos que comandam o que é apresentado às pessoas nas plataformas de redes sociais, a ponto de impactar seus resultados. Alguns aspectos da interseção desse problema com o jornalismo podem ser encontrados em Nechushtai e Lewis (2019), Sánchez e Ruiz (2020) e Saquete et al. (2020).

Se a produção do que se convencionou chamar de *fake news*, no sentido de desinformação, não é essencialmente um expediente novo na história da humanidade, o ecossistema informacional contemporâneo lhe concedeu escala, potencial de impacto e dificuldade de resolução até então inéditos.

O recente crescimento das iniciativas e órgãos de checagem de fatos e combate às notícias falsas, além de fundamental, abriu um caminho ainda tímido para a defesa da necessidade do jornalismo tradicional e seu processo de mediação e depuração. Crises mundiais como a pandemia do coronavírus têm oferecido ao produto jornalístico tradicional uma retomada, pelo menos parcial, da centralidade e alcance de tempos anteriores.

Diante do quadro geral apresentado e considerando o recorte proposto inicialmente, ou seja, que vai além do olhar exclusivo sobre o processo de produção do conteúdo jornalístico, ainda que em caráter exploratório, que outras áreas de ação poderiam ser ativadas, de forma a colaborar com melhores resultados da indústria jornalística e aumento da percepção da relevância do seu produto diante de seus consumidores?

3 Dados e algoritmos na reconfiguração do contato com o usuário

As tecnologias emergentes citadas anteriormente, que em outras indústrias são consideradas de alto valor estratégico para operação, na indústria do jornalismo são consideradas, em geral, apenas para o processamento das informações que serão convertidas em conteúdo jornalístico e não para uso sobre o conjunto de informações gerado pelo relacionamento diário dos usuários com esse produto.

Um conjunto de possibilidades que listaremos a seguir permanece subutilizado pelo segmento da produção de conteúdo informativo e representa um potencial de aproximação mais personalizada e efetiva com os usuários, a partir dos dados que estes mesmos geram na interação com as publicações.

Dentro da grande área da inteligência artificial encontramos segmentos como o aprendizado de máquina (*machine learning* – ML), a visão computacional e os estudos de processamento de linguagem natural, entre outros.

Grus (2016, p.142) propõe uma definição do termo aprendizado de máquina “para nos referir à criação e uso de modelos que são aprendidos a partir de dados”. Assim, a partir de dados preexistentes os modelos de ML tentam “prever possíveis saídas para dados novos”.

A utilização desse tipo de abordagem é muito comum, por exemplo, para verificar se determinado e-mail que recebemos é spam, por exemplo. Dentre as principais tarefas ligadas ao aprendizado de máquina, Amaral (2016, p.7) define “classificação, regressão, agrupamentos e regras de associação”.

Na classificação, diante de um elemento novo ou não identificado, o objetivo é, a partir de um conjunto pré-definido de grupos de elementos que têm atributos em comum, denominados classes, definir a qual classe esse elemento novo pertence. Segundo Amaral (2016, p.7), “usamos classificação para prever uma fraude, descobrir a qual espécie um animal pertence, prever uma doença ou classificar um tipo de fungo”.

A regressão, ao invés de trabalhar com categorias ou dados nominais, trabalha com dados numéricos e pode ser utilizada, por exemplo, para prever o peso de uma pessoa a partir do número de calorias que ela ingere diariamente.

No agrupamento ou clusterização, não existem classes pré-existentes e começamos com vários elementos analisando seus

atributos para, a partir deles, propor grupos ou *clusters* de elementos que compartilham atributos semelhantes e daí dividi-los em grupos.

Dependendo do tipo de agrupamento que utilizarmos, um elemento pode pertencer a mais de um grupo ou não ser agrupado, ou seja, ser considerado ruído. Exemplos de uso: identificar grupos de clientes e direcionar campanhas de marketing específicas, identificar tentativas de acesso à rede; categorizar uma nova espécie entre outros. (Amaral, 2016, p.8).

Por fim, as regras de associação tentam identificar relações entre itens. Um exemplo bem comum é a prática no *e-commerce* de associar compras entre usuários, ou seja, a partir dos dados de quem comprou o item A, comprou também itens B e C, tal informação ser fornecida a novos consumidores como sugestão.

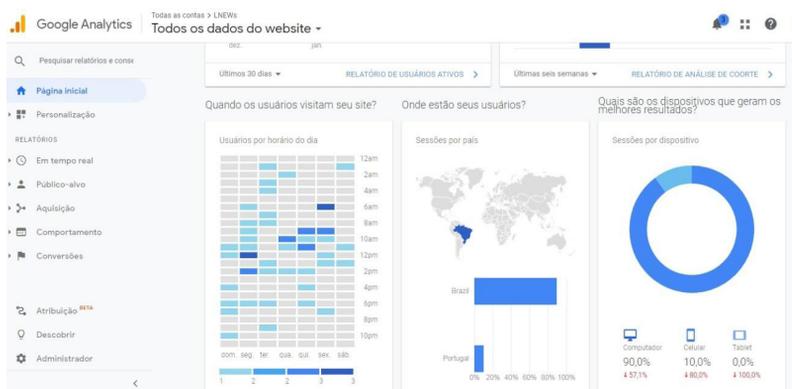
Uma série de tipos de algoritmos está associada à implementação de modelos de ML para suas diversas aplicações. *Bayes*, Frequência de Subconjuntos e Árvore de Dados são alguns (Amaral, 2016; Grus, 2016), e seu detalhamento vai além do escopo deste texto.

As aplicações que discutiremos a seguir, em sua maioria, estão relacionadas ao aprendizado de máquina e sua utilização na reconfiguração do relacionamento das empresas jornalísticas com seus usuários. A escolha por *Google Analytics* deve-se ao fato de ser a solução gratuita e de maior adoção para geração de métricas no ambiente web. Já *NewsWhip*³ representa um exemplo de startup que conseguiu integrar nas suas soluções o tipo de análise preditiva que descrevemos aqui, tendo a vantagem de disponibilizar uma extensão para o browser Chrome que pode ser instalada para testes por pesquisadores sem nenhum custo.

3.1 Identificação de padrões de interesses

A identificação de padrões é uma das possibilidades mais comuns entre as diversas aplicações da inteligência artificial. Hoje, com ferramentas como *Google Analytics*, é possível coletar dados diversos sobre o comportamento do usuário na interação com o produto jornalístico. As páginas que visitou, o tempo dedicado a cada uma, a origem da conexão, o tipo de dispositivo usado para acessar o conteúdo, o sistema operacional e a localização geográfica do usuário.

Figura 1 – Exemplo de tela de dados do *Google Analytics*



Fonte: do autor

Esse conjunto de dados pode ser utilizado para, através das operações tradicionais das ferramentas de ML, propor conteúdos, associar produtos a partir dos perfis de consumo e notificar o usuário para a publicação de notícias nas categorias de maior interesse.

É importante ressaltar que essas possibilidades complementam um papel antes apenas dedicado aos editores que é decidir, com base na sua própria experiência profissional, o que destacar ou cobrir com mais profundidade entre os diversos fatos do dia a dia.

A disponibilidade desse pacote de dados adicionais sobre o comportamento dos consumidores de informação permite um nível de personalização antes impossível de atingir a partir de métricas mais genéricas como circulação e audiência, ambas oriundas do ambiente analógico e difusas demais para permitir um nível de granularidade individualizado.

3.2 Predição de Engajamento Social

Os modelos de predição de tendências podem ser aplicados ao ambiente do jornalismo, expandindo o alcance das diversas soluções de monitoramento de plataformas de mídias sociais, com utilização hoje comum em vários segmentos empresariais.

O que o aprendizado de máquina pode fazer é, a partir de um conjunto de dados pré-existente, consolidado usando os relatórios de monitoramento, avaliar e prever qual o nível de engajamento futuro

que determinada notícia vai ter, ou seja, se determinada informação tem maior ou menor chance de gerar interações crescentes com o público daquele canal de mídia.

Empresas como *NewsWhip* têm como atividade principal desenvolver iniciativas baseadas em ML para otimizar os processos de produção de jornalismo. Por intermédio de vários produtos a empresa oferece às redações a possibilidade de avaliar antecipadamente, por exemplo, se um *tweet* tem potencial de viralização, dando ao jornalista um alerta para possível furo.

Utilizando o termo “inteligência preditiva de mídia” e métricas como “*social velocity*”, que medem a proporção de engajamento ao longo do tempo, a empresa utiliza algoritmos de regressão para calcular a estimativa de novas interações a partir da publicação de determinada matéria como no exemplo abaixo.

Na matéria que utilizamos do *Estadão.com*, que relata as recomendações do MPF para o governo sobre o coronavírus, a ferramenta de *NewsWhip* para o Chrome permitiu não só identificar 644 interações com a matéria no *Facebook*, como o algoritmo de monitoramento tradicional faria, como previu 262 novas interações nos próximos trinta minutos.

Figura 2 – Exemplo de tela de uma das soluções de *NewsWhip* prevendo o engajamento futuro



Fonte: do autor

Esse tipo de resultado, que pode ser agregado em painéis e visualizado em conjunto com o de outras matérias, oferece uma nova camada de informação para a tomada de decisões na redação.

3.3 Alocação de recursos para coberturas

Além dos dados, até em tempo real, oferecidos pelas soluções de *analytics* e dos registros e séries históricas de interesse, captados por ferramentas como o *Google Trends*⁴, já bastante conhecidas, as ferramentas que usam aprendizado de máquina podem impactar até questões logísticas e de alocação de recursos humanos ou materiais para dar, por exemplo, mais atenção e espaço para tópicos com potencial maior de crescimento nas métricas de engajamento.

Um tópico com potencial ascendente pode ser passado a uma equipe especializada ou ser encaminhado para análise de especialistas, agilizando o processo normal de gerar conteúdo, que vai repercutir e aprofundar a cobertura factual mais imediata.

Mesmo recursos como carros, motos, links de internet móvel e voos de helicóptero podem ser otimizados a partir de modelos preditivos que podem cruzar não apenas as rotas de cobertura das notícias em ascensão como as condições de trânsito, custos envolvidos e pessoal disponível naquele dia ou plantão. O impacto disso na gestão financeira da operação jornalística pode ser significativo.

A alocação do próprio conteúdo, na estruturação das páginas que o usuário vai acessar, pode, em tese, sofrer significativo aumento do nível de personalização a partir da combinação das diversas possibilidades listadas anteriormente.

Hoje, a partir da lógica bastante explorada das bases de dados, é comum, entre os portais de notícia, gerar uma página personalizada apenas com o conteúdo de textos, fotos e vídeos sobre determinado time de futebol, por exemplo, utilizando técnicas simples como o uso de *tags* (rótulos), palavras-chave e categorias.

A partir das soluções de aprendizado de máquina é possível, sem grandes custos adicionais, ordenar tal conteúdo de acordo com as preferências de um torcedor específico, a partir dos dados gerados por ele em suas interações anteriores. Exemplificando, se o perfil dele indica o consumo preferencial de vídeos do seu time de coração, a página apresentada terá esses itens em destaque ou na parte de

cima, sem necessidade de rolagem, para que ele acesse de forma mais imediata aquilo que gosta mais.

Obviamente tal tipo de serviço estaria disponível apenas para assinantes que depois de fazer seu *log in* seriam identificados, possibilitando o carregamento do seu perfil de utilização e consequentemente a adaptação do conteúdo e de sua forma de apresentação aos seus interesses. Este é um bom exemplo do que chamamos de personalização estendida oportunizando a monetização, via venda de assinatura digital, potencializada por diferenciais de serviços e conteúdos exclusivos, incluindo material imersivo em realidade virtual ou agregado via realidade aumentada como descrito em Santos (2015).

4 Considerações Finais

No presente texto, discute-se a reconfiguração estratégica da utilização de dados dentro das organizações jornalísticas, por meio de algoritmos e soluções de inteligência artificial, entre elas, especificamente, o aprendizado de máquina como alternativa para elevação da percepção de valor sobre o produto informativo, reduzida pela fragmentação das audiências, explosão de emissores, concorrência indireta das grandes empresas de tecnologia e transformações culturais do ecossistema de meios digitais.

Tal abordagem propõe a identificação de padrões de interesse, a predição de engajamento social e a alocação de recursos para coberturas como formas, entre outras possíveis, de expandir o atual nível de personalização oferecido aos consumidores de notícia.

A utilização de soluções de inteligência artificial e especificamente de aprendizado de máquina não é perfeita, nem isenta de dificuldades e riscos.

Estudos recentes de Amadeu (2019) e Silva (2019), por exemplo, têm alertado para situações de invasão da privacidade, distorção dos fatos e práticas como racismo e xenofobia, inseridas nos resultados ou proposições definidas a partir de algoritmos de ML.

Como sequências de instruções conhecidas em detalhes apenas pelos programadores, as soluções de inteligência artificial e seus algoritmos, especificamente, aprendizagem de máquina, normalmente constituem-se como caixas pouco transparentes, de difícil entendimento para leigos, tanto em termos das funcionalidades oferecidas, bem como de riscos ou vieses internamente presentes.

Inevitáveis e onipresentes, as soluções de código fazem parte do suporte operacional de quase todas as atividades humanas contemporâneas, representando ao mesmo tempo a beleza e a mágica dos primeiros autômatos (Devaux, 1964) até descontrolo e riscos (Ellul, 1968) reais ou ficcionais sobre os quais tantos autores, pesquisadores e roteiristas de cinema se debruçaram. Trabalhos de filósofos da tecnologia como Feenberg (2002, 2010) tentam equalizar prós e contras de um processo contemporâneo.

A utilização estratégica dos dados via algoritmos é o modelo de negócio das principais empresas globais de tecnologia como *Google* e *Facebook*, em parte, responsáveis pela redução da percepção de valor do produto jornalístico tradicional, inclusive por sua utilização sem custos em suas ferramentas de busca e plataformas de mídias sociais respectivamente.

Como grandes concentradoras de dados, não apenas das notícias, mas sobre o entorno social onde operam, as empresas de mídia foram obrigadas a portar seus conteúdos, produzidos com recursos próprios, para os canais digitais de difusão de informação controlados pelas empresas de tecnologia, como forma de minimizar a fragmentação das audiências que elas mesmas ajudaram a criar.

Nessa situação, a valorização dos dados de que dispõem torna-se um caminho de potencial estratégico que deve ser considerado. Sua utilização não só nos aspectos relatados neste artigo (identificação de padrões de interesse; predição de engajamento social; alocação de recursos para coberturas) é bem mais ampla.

Entendemos que, para enfrentar o problema central da redução da percepção de valor do produto jornalístico tradicional pela sociedade, a utilização de algoritmos, dados e tecnologias emergentes, com efetividade e nível ampliado em termos de personalização, são um caminho a ser explorado.

O combate às *fake news* e a momentânea centralidade recuperada por conta da pandemia não nos parecem ser suficientes para resolver questões estruturais e anteriores.

Destaque-se aqui a proposição de caráter apenas exploratório deste texto, longe de estabelecer respostas definitivas, tendo em vista a complexa rede de possibilidade e imbricações entre o fazer jornalístico e a tecnologia.

NOTAS

- 1 Dados que podem ser acessados e utilizados, sem restrições ou custos, inclusive comercialmente, no máximo exigindo apenas o crédito da fonte.
- 2 No original: “Computational Journalism is the application of computing and computational thinking to the activities of journalism including information gathering, organization and sensemaking, communication and presentation, and dissemination and public response to news information”.
- 3 www.newswhip.com/
- 4 trends.google.com.br/

REFERÊNCIAS

- Amadeu, S. (2019). *Democracia e os códigos invisíveis: Como os algoritmos estão modulando comportamentos e escolhas políticas*. São Paulo: Edições Sesc (Coleção Democracia Digital).
- Amaral, F. (2016). *Aprenda mineração de dados – Teoria e prática*. Rio de Janeiro: Alta Books.
- Barbosa, S. (2007). *Jornalismo Digital em Base de Dados (JDBD) – Um paradigma para produtos jornalísticos digitais dinâmicos* [tese de doutorado, Universidade Federal da Bahia]. Repositório Institucional UFBA.
- Barbosa, S. (2008). Modelo Jornalismo Digital em Base de Dados (JDBD) em Interação com a Convergência Jornalística. *Revista de la Sociedad Española de Periodística*, (1), 87-106.
- Barbosa, S. (2009). Modelo JDBD e o ciberjornalismo de quarta geração. In J. M. Flores Vivar, & F. E. Ramírez (Orgs.), *Periodismo Web 2.0* (pp.271 – 283). Madrid, ES: Editorial Fragua.
- Barbosa, S. (2011). Jornalismo em ambientes dinâmicos: Perspectivas, tendências e desafios para a criação de conteúdos em tempos de convergência. In A. L. Ureta, & K. M. Ayerdi (Orgs.), *III Congreso Internacional de Ciberperiodismo y Web 2.0* (pp.7 – 17). Bilbao, Espanha: Universidad del País Vasco.

Barbosa, S., & Torres, V. (2013). O paradigma 'Jornalismo Digital em Base de Dados': Modos de narrar, formatos e visualização para conteúdos. *Galaxia*, (25), 152-164. DOI: 10.1590/S1982-25532013000200013

Carlson, M. (2014). The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms and journalistic authority. *Digital Journalism*, 3 (3), 416-431. DOI: 10.1080/21670811.2014.976412

Carreira, K. (2017). *Notícias Automatizadas – A evolução que levou o jornalismo a ser feito por não humanos* [dissertação de mestrado, Universidade Metodista]. Biblioteca digital de teses e dissertações.

Clerwall, C. (2014). Enter the Robot Journalist: User's perception of automated content. *Journalism Practice*, 8 (5), 519 – 531. DOI: 10.1080/17512786.2014.883116

Coddington, M. (2014). Clarifying Journalism's Quantitative Turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3 (3), 331–348. DOI: 10.1080/21670811.2014.976400

Coppin, B. (2010). *Inteligência artificial*. Rio de Janeiro: LTC.

Dalen, A. (2012). The Algorithms Behind the Headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6 (5-6), 648-658. DOI: 10.1080/17512786.2012.667268

Devaux, P. (1964). *Autômatos, automatismo e automatização*. Lisboa: Editorial Gleba.

Diakopoulos, N. (2011, abril). *A Functional Roadmap for Innovation in Computational Journalism*. Nicholas Diakopoulos – Musings on Media. Recuperado de www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/

Ellul, J. (1968). *A técnica e o desafio do século*. Rio de Janeiro: Paz e Terra.

Feenberg, A. (2002). *Transforming technology: A critical theory revisited*. New York: Oxford University Press.

Feenberg, A. (2010). *Between reason and experience. Essays in technology and modernity*. Cambridge, MA: Mit Press.

Fidalgo, A. (2007). A resolução semântica no jornalismo online. In S. Barbosa (Org.), *Jornalismo digital de terceira geração* (pp.93 – 102). Covilhã, PT: Livros LabCOM.

Fidalgo, A. (2004). Do poliedro à esfera: os campos de classificação. A resolução semântica no jornalismo online. Anais do II Encontro Nacional da SBPJor. Salvador, BA.

Graefe, A. (2016, jan.). *Guide to Automated Journalism*. Tow Center for Digital Journalism and James S. and John L. Knight Foundation. Recuperado de www.gitbook.com/book/towcenter/guide-to-automated-journalism/details

Grus, J. (2016). *Data Science do Zero*. Rio de Janeiro: Alta Books.

Lage, N. (1997). O lead clássico como base para a automação do discurso informativo. Anais do 20 Congresso Brasileiro de Pesquisadores da Comunicação Intercom. Santos, SP.

Latar, N. (2015). The Robot Journalism in the Age of Social Physics: The end of human journalism? In G. Einav (Org.), *The New World of Transitioned Media* (pp.65 – 80). Springer. DOI: 10.1007/978-3-319-09009-2_6

Lewis, S. & Usher, N. (2014). Collaboration and The Future of Journalism: A case study of the Hacks/Hackers global network. *Digital Journalism*, 2 (3), 383-393. DOI: 10.1080/21670811.2014.895504

Machado, E. (2006). *O Jornalismo Digital em Base de Dados*. Florianópolis, SC: Calandra.

Maier, S. R. (2002). Numbers in the News: A Mathematics Audit of a Daily Newspaper. *Journalism Studies*, 3 (4), 507-519. DOI: 10.1080/1461670022000019191

Meyer, P. (1973). *Precision Journalism. A reporter's introduction to social science methods*. Bloomington: Indiana University Press.

Nechushtai, E., & Lewis, S. (2019). What kind of news gatekeepers do we want machines to be? filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in human behavior*, 298–307. DOI: 10.1016/J.CHB.2018.07.043

Rubio, L., & Ruiz, M. (2020). Percepción de docentes universitarios, estudiantes, responsables de innovación y periodistas sobre el uso de inteligencia artificial en periodismo. *El profesional de la información*, 29(1), 1–14. DOI: 10.3145/EPI.2020.ENE.09

Sánchez, J., & Ruiz, M. (2020). Inteligencia artificial y periodismo: una herramienta contra la desinformación. *Revista Cidob d'afers internacionals*, (124), 49–72. DOI: 10.24241/RCAI.2020.124.1.49

Santos, M. (2014). Textos Gerados por Software – Surge Um Novo Gênero Jornalístico? Anais do XXXVII Congresso Brasileiro de Ciências

da Comunicação. Foz do Iguaçu, Paraná. Recuperado de www.intercom.org.br/papers/nacionais/2014/resumos/R9-2066-1.pdf

Santos, M. (2015). Jornalismo, mobilidade e realidade aumentada: notas sobre possibilidades de utilização. In J. Canavilhas, & I. Satuf (Orgs.), *Jornalismo para Dispositivos Móveis: Produção, distribuição e consumo* (pp.103 –126). Covilhã, Portugal: UBI.

Santos, M. (2016a). Narrativas Automatizadas e a Geração de Textos Jornalísticos: A estrutura de organização do lead traduzida em código. *Brazilian Journalism Research*, 12(1), 160-185. DOI: 10.25200/BJR.v12n1.2016.757

Santos, M. (2016b). Internet das Coisas e Sistemas Inteligentes no Jornalismo. *Comunicação e Inovação*, 17 (34), 21-39. DOI: 10.13037/ci.vol17n34.3769

Santos, M. (2016c). Jornalismo e Internet Das Coisas – Notas sobre tipologia e modelos de utilização. *Revista Latinoamericana de Ciencias de La Comunicación*, 12 (22), 134–145. Recuperado de www.alaic.org/revista/index.php/alaic/article/view/621/380

Santos, M. (2019a). Narrativas e desenvolvimento de conteúdo imersivo: Aplicação de sistemas bi e multissensoriais de realidade virtual no jornalismo. *Revista da Intercom*, 42 (3), 133-150. DOI: 10.1590/1809-5844201937

Santos, M. (2019b). A datificação de um campo de conhecimento. *Organicom USP*, (31), 145–157. Recuperado de www.revistas.usp.br/organicom/article/view/161444/159283

Saquete, E., Tomás, D., Moreda, P., Martínez-Barco, P., & Palomar, M. (2020). Fighting post-truth using natural language processing: a review and open challenges. *Expert systems with applications*. DOI: 10.1016/J.ESWA.2019.112943

Silva, T. (2019). Linha do tempo do racismo algorítmico. *Blog do Tarcízio Silva*. Recuperado de <https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo>

López, J. M., Bran, C., & Requeijo, S. (2018). Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en España. *El profesional de la información*, 27 (4), 750–758. DOI: 10.3145/EPI.2018.JUL.04

Višnovský, J., Ungerová, M., & Kubíková, K. (2019). Robo-journalism and its implementation in editorial practice. *Anais do NORDSCI Conference*. Recuperado de <https://files.eric.ed.gov/fulltext/ED603467.pdf>

MÁRCIO CARNEIRO DOS SANTOS. Doutor pelo programa de Tecnologias da Inteligência e Design Digital (TIDD) da PUC-SP. Professor permanente dos programas de Pós Graduação em Design e Pós Graduação em Comunicação – Mestrado Profissional, ambos na UFMA. Professor do Departamento de Comunicação Social na área de Jornalismo em Redes Digitais. Coordenador do LABCOM – Laboratório de Convergência de Mídias. Líder do grupo de pesquisa CNPq – Tecnologia e Narrativas Digitais – TECND. E-mail: mcszen@gmail.com

PARECERES

Um dos pareceres utilizados na avaliação deste artigo pode ser acessado em: <https://osf.io/gcr3z/>

Seguindo a política de ciência aberta da BJR, o avaliador autorizou a publicação do parecer e a divulgação do seu nome.