



BIG DATA E A TRANSFORMAÇÃO DA VIDA SOCIAL E POLÍTICA CONTEMPORÂNEA

O nosso mundo é hoje moldado por uma tecnologia que se impôs e que condiciona as atividades sociais, económicas, políticas e culturais. O atual relevo da *big data* e o caminho para a “dataficação” constituem revoluções impactantes com enormes consequências na vida coletiva, às quais o poder político tem de dar respostas.

Afirmar que o nosso tempo é o resultante de uma enorme mudança, em grande medida continuada pela Revolução Industrial, parece ser uma evidência. Resta perguntar: para onde? Quem será protagonista? Estaremos em condições de antecipar as suas reais consequências? O nosso quotidiano está há algum tempo a ser moldado por uma tecnologia que se impôs e que condiciona as atividades sociais, económicas, políticas e culturais. A Inteligência Artificial (IA) e a Internet das Coisas (*Internet of Things* – IoT) são apenas dois exemplos de áreas que, nas últimas décadas, têm sido alvo de investigação, quer aplicada quer ao nível dos seus efeitos.

Se aos observadores, consumidores e também cidadãos parece estar a vislumbrar-se uma transformação motivada por tão impactantes tecnologias, é por isso urgente identificar alguns dos seus efeitos, reconhecendo que o que está a desenvolver-se carece de ponderação e análise, mais necessária num tempo tão acelerado. Deste modo, ensaiamos neste texto algumas consequências sociais e políticas do peso dos dados na vida coletiva e de uma nova forma de fazer política.

Relevo atual da *big data*

Todos nós somos produtores de dados e, de acordo com o entendimento hodierno, somos criadores de uma riqueza tal que somos alvos especialmente fáceis das grandes empresas de tecnologia (*big techs*). Como utilizadores da Internet e como consumidores, deixamos uma pegada digital facilmente rastreável. Se uma parte importante das nossas vidas já decorre *online*, condição para não nos desligarmos sem mais, haverá motivo para admiração? *Big data* é um tema central na abordagem aos dados, processamento e tratamento da informação em larga escala. Se parece um chavão dizer que “os dados são o ouro do século XXI”, tese que Harari (2017) des-

monta, essa ideia encerra em si mesmo um aspeto decisivo: as empresas, os negócios e outras organizações só conseguirão ser verdadeiramente competitivas e capazes de vingar se reunirem, analisarem e visualizarem a informação das múltiplas variáveis dos respetivos bens, produtos e serviços onde operam, e assim responderem de modo mais eficaz e rápido às necessidades de cada momento (Harari, 2024).

... *Big data* é uma expressão que condensa diferentes conjuntos de dados em grande número, altamente complexos, não-estruturados, organizados, armazenados e processados, recorrendo a métodos e técnicas específicas usadas para processos de negócios.

De acordo com De Mauro *et al.* (2015), a expressão *big data* é um modismo tanto na indústria como na academia e tem sido apresentada de uma maneira muito vaga. Com efeito, não traduz algo completamente novo, ainda que tenda a descrever um conjunto de situações que se estão a consolidar, que vão desde a capacidade tecnológica para armazenar, concentrar e processar dados, até à mudança cultural que está a conquistar o mundo empresarial e a sociedade, marcada por uma grande abundância de informação. Os autores mencionam que a grande maioria das referências a *big data* englobam-se num destes quatro temas: informação, tecnologias, métodos e impacto.

Se a informação é o combustível da *big data*, também é certo que, por via da digitalização, tem sido mais fácil aceder-lhe e em muitos ocasiões manipulá-la. Um próximo nível é a *dataficação*, isto é, garantir um formato quantificável da informação, para que a mesma possa ser tabulada e analisada. No que respeita às tecnologias, remete-se para os equipamentos e a arquitetura de sistemas que permitem a transmissão de grandes quantidades de dados e a monitorização de desempenho dos mesmos; no que concerne aos métodos, a análise de grande volume de dados exige métodos de processamento que vão além das ferramentas estatísticas tradicionais. Podem destacar-se, a título de exemplo, vários métodos de análise de dados da *big data*, dentre eles o processamento de linguagem natural, modelos de previsão e algoritmos genéticos; relativamente ao impacto da *big data* nas nossas vidas, um elemento fundamental é o da privacidade, ou seja, a garantia que temos (ou não) que os dados que partilhamos ou disponibilizamos sejam circunscritos apenas para determinados fins. Esse é, precisamente, um dos principais fins da recente Diretiva comunitária da proteção de dados.

Num esforço para definir o conceito, Ularu *et al.* (2012) consideram que *big data* é uma expressão que condensa diferentes conjuntos de dados em grande número, altamente complexos, não-estruturados, organizados, armazenados e processados, recorrendo a métodos e técnicas específicas usadas para processos de negócios. Os dados excedem a capacidade de processamento dos sistemas de base de dados convencionais. Por outras palavras, a sua dimensão e velocidade de circulação impedem que se encaixem em algumas estruturas existentes, daí a necessidade de desenvolver novas arquiteturas.

O enquadramento e as propostas de sistematização da *big data* levam a destacar os “três V”: volume, velocidade e variedade.

A quantidade de dados e a sua natureza descreve essa perspectiva do volume, que pode variar de dezenas de terabytes para centenas de petabytes para uma organização. Ao recuperarem as previsões para o ano de 2020, Santaella e Kaufman (2021) indicam que as informações do mundo dobram a cada dois anos, pelo que, em 2020, estaríamos na ordem de quarenta ZB (zetabytes). A previsão para 2025 é de um volume de dados de 175ZB, crescimento deveras significativo, considerando que em 2010 foram produzidos “somente” dois ZB.

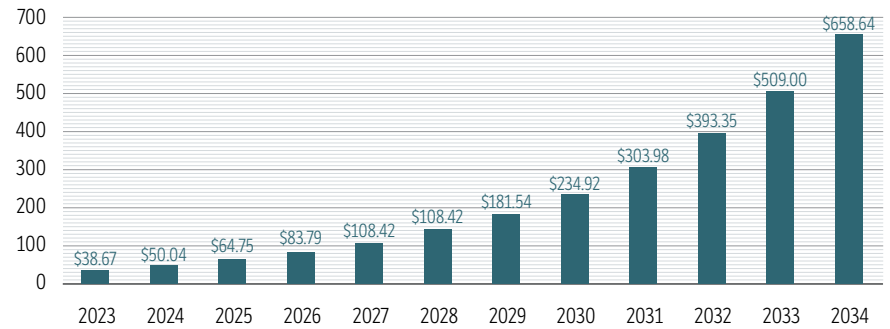
A expressão “tsunami de dados” adquiriu especial visibilidade a partir de 2014, motivada especialmente pela proliferação de redes sociais, juntamente com um acréscimo de dispositivos móveis e em rede (Internet das Coisas), finanças e retalho *online*, além de avanços no setor físico e setores de ciências da vida. Em 2014, o então Twitter já processava aproximadamente doze TB (terabytes) de dados por dia, enquanto o Facebook recebia mais de quinhentos milhões de *likes* diários. Em 2020, a Google processa, em média, mais de 40.000 consultas de pesquisa a cada segundo, o que perfaz mais de 3,5 bilhões de pesquisas por dia e 1,2 trilhões de pesquisas por ano. De acordo com a IBM, a cada segundo, 300 horas de vídeo são carregadas no YouTube, 204 milhões de e-mails são enviados e 500.000 comentários são postados no Facebook. As autoras sublinham que, para 2020, já teríamos alcançado 50 bilhões de dispositivos conectados à Internet e que 1,7 bilhões de novas informações seriam desenvolvidas, por segundo, por todos os habitantes ao redor do mundo (Santaella & Kaufman, 2021: 216). A referência ao volume de dados criados no mundo provém de inúmeras fontes que não exclusivamente os *posts* nas redes sociais, mas igualmente notícias, vídeos e fotografias digitais, registros de compras, transações financeiras, sinais de GPS de telefones móveis, e-mails, informação meteorológica, entre outros.

A velocidade remete para os dados que uma organização recebe e que precisa ser trabalhada em tempo real. Já a variedade de dados não estruturados, estruturados e semiestruturados afeta o modo como as organizações podem sumariar e analisar a sua informação.

Mark Weiner e Sarab Kochhar (2016) acrescentam a ação de outro “V”: valor. Querem os autores com isto chamar a atenção para o valor intrínseco dos dados que ajudam as

FIGURA 1. DIMENSÃO DO MERCADO DE ANÁLISE DE DADOS DE 2023 A 2034
(em milhares de milhão de USD)

Fonte: <https://www.precedenceresearch.com/data-analytics-market>



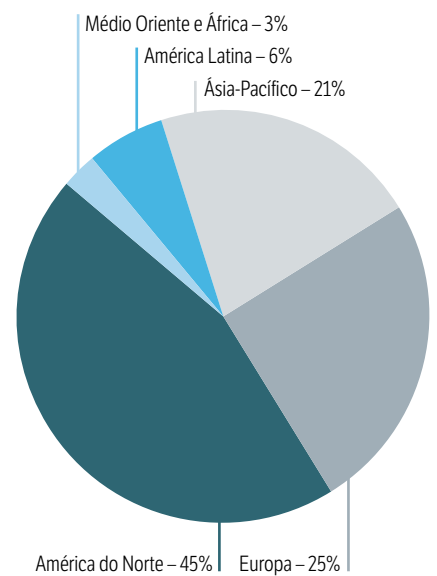
instituições a obter significado, reconhecer padrões e a fazer opções mais esclarecidas no processo de tomada de decisão. As terminologias podem até mudar nos próximos anos, mas a necessidade de criar estratégias, recolher e analisar os dados irão tornar-se uma prioridade de topo para as organizações.

A Comissão Europeia (2020) sublinha que cada nova vaga de dados oferece à União Europeia (UE) grandes oportunidades para se tornar líder mundial neste segmento. A Comissão antecipa uma mudança radical no modo de armazenamento e de tratamento dos dados nos próximos anos. Estima-se que 80% do tratamento e da análise de dados ocorrem em centros de dados e instalações de computação centralizadas e 20% em objetos inteligentes conectados – designadamente automóveis, eletrodomésticos ou robôs de fabrico – e em instalações de computação próximas do utilizador (“computação periférica”). É expectável que estas proporções se invertam até 2025. Para além de vantagens em termos económicos e de sustentabilidade, este desenvolvimento cria novas oportunidades para as empresas gerarem instrumentos que permitam aos produtores de dados aumentar o controlo sobre os mesmos.

Reconhece-se que a UE tem potencial para ser bem-sucedida na economia ágil dos dados. Tem ao seu dispor tecnologia, conhecimentos especializados e mão-de-obra altamente qualificada. Porém, concorrentes como a China e os Estados Unidos da América (EUA) avançam rapidamente no acesso e utilização de dados em todo o mundo. Nos EUA, a organização do espaço de dados é deixada ao setor privado, o que produz efeitos de concentração assinaláveis. A China conjuga a vigilância governamental com um controlo vigoroso das grandes empresas tecnológicas sobre

FIGURA 2. QUOTA DE MERCADO DE ANÁLISE DE DADOS, 2023
(por região – %)

Fonte: <https://www.precedenceresearch.com/data-analytics-market>



quantidades gigantescas de dados, sem fornecer garantias suficientes às pessoas (Comissão Europeia, 2020).

O caminho para a *dataficação*

O comércio dos dados é planetário e o volume financeiro dessas transações está a crescer exponencialmente. Contudo, falta regulação e é preciso antecipar impactos da venalidade dos dados.

Bezerra (2024) sustenta que, ao denunciar-se a ideologia dos dados que está por trás do endeusamento da tecnologia nos nossos dias, está a defender-se a soberania tecnológica contra os atuais expedientes do colonialismo de dados, uma nova forma de apropriação da vida humana exercida pelas práticas de extração de dados para fins lucrativos, e uma realidade distópica em que a vida humana é não só anexada ao capitalismo, mas simultaneamente sujeita a monitorização e vigilância permanentes.



Van Dijck (2014, p. 198) identifica duas características centrais no que define “ideologia do dataísmo”: a primeira é a crença generalizada na quantificação objetiva e no rastreamento potencial de todos os tipos de comportamento humano e sociabilidade por meio de tecnologias de *media online*; a segunda remete para a confiança nos agentes (institucionais) que reúnem, interpretam e compartilham metadados selecionados de redes sociais, plataformas de Internet e outras tecnologias de comunicação. Deste modo, para a autora, a *dataficação* tornou-se um novo paradigma aceite para entender a sociabilidade e o comportamento social. Soma-se neste espectro ideológico a noção de inevitabilidade, de que fala Zuboff (2020), como se o desenvolvimento tecnológico e a aplicação de novas tecnologias só pudessem ocorrer sob a égide de grandes empresas capitalistas.

A identificação de padrões de conduta ou atividades a partir de metadados deixados de modo inconsciente em *sites* de redes sociais serve, cada vez mais, para prever comportamentos futuros. A monitorização de redes sociais fornece informações significativas para a polícia e os serviços de *intelligence* preverem atividades terroristas emergentes ou calcular o controlo de multidões, bem como para os profissionais de marketing preverem preços futuros da bolsa ou receitas potenciais de bilheteira. Do ponto de vista da vigilância e do marketing, a analítica preditiva gera informações bastante relevantes sobre quem somos e o que fazemos. No que respeita ao comportamento humano, essa lógica poderá tornar-se intrigante, resvalando para um terreno ora de análise e projeção, ora de dedução e previsão (van Dijck, 2014).

Ainda que se acredite que os metadados extraídos das plataformas de redes sociais reflitam o comportamento humano como ele é, os algoritmos empregados pela Google, Twitter (hoje conhecido como X) e quejandos são intrinsecamente seletivos e manipuladores. Ao se preverem preferências políticas analisando o desempenho do debate por meio de *tweets*, não devemos ignorar o potencial que outros técnicos, como *spin-doctors*, ou *twitteiros* partidários têm de influenciar os debates na rede X, em tempo real. No domínio do marketing, a previsão das necessidades dos futuros clientes é semelhante à manipulação do desejo: detetar padrões específicos nos hábitos do consumidor acaba

por resultar em tentativas simultâneas de criar procura – uma estratégia de marketing que é monetizada com sucesso pelo inefável algoritmo de recomendação da Amazon (van Dijck, 2014).

O dataísmo presume confiança na objetividade dos métodos quantificáveis, bem como na independência e integridade das instituições que implantam esses métodos, quer sejam plataformas corporativas, agências governamentais ou académicos. Num ecossistema de conectividade, confiança e independência são noções em conflito, reconhecendo-se que todas as plataformas *online* estão inevitavelmente interconectadas, seja ao nível da infraestrutura seja ao nível de lógica operacional, algo que nos aproxima do pensamento de Michel Foucault muito antes da emergência de um mundo digital (Lorenzetto & Teixeira Filho, 2022).

... a capacidade dos utilizadores de resistir às políticas de privacidade e táticas de vigilância das plataformas é muito limitada.

A divulgação de táticas rotineiras de vigilância de dados, tal como denunciadas com estrondo por Edward Snowden quando era funcionário da National Security Agency (NSA), ameaçou minar irremediavelmente não só a confiança das pessoas em agências estatais ou empresas individuais, mas igualmente os fundamentos institucionais do dataísmo como um todo. Van Dijck (2014) dá nota que a monitorização dos cidadãos com base em seus dados *online*, conhecido como *dataveillance*, difere da vigilância em pelo menos um ponto importante – se a vigilância pressupõe monitorização para fins específicos, já a vigilância dos dados envolve o rastreamento contínuo de metadados para fins predefinidos não declarados. A vigilância dos dados introduz-se, pois, em todo o domínio social. *Dataveillance* é, com efeito, uma proposta de longo alcance, portadora de profundas consequências para o contrato social entre, de um lado, plataformas corporativas e agências governamentais e, do outro, cidadãos-consumidores. A autora sublinha, que nos últimos anos, o poder real dos utiliza-

dores-cidadãos em relação às plataformas corporativas e ao Estado espoletou um debate aceso principalmente entre ativistas e académicos. Descobriu-se que a capacidade dos utilizadores de resistir às políticas de privacidade e táticas de vigilância das plataformas é muito limitada. Os indivíduos são conduzidos pelas tecnologias das plataformas e pelos modelos de negócios de plataformas únicas, ainda que seja muito difícil obter informações sobre a interdependência e a complexidade do sistema; argumentou-se também a favor do fortalecimento da alfabetização digital (consumidor), especificamente no nível de compreensão da privacidade e segurança em relação aos dados sociais (van Dijck, 2014, pp. 204-207). Outros há que enfatizam a necessidade de os utilizadores descobrirem como os *media* digitais forjam um novo contrato social nas sociedades, ao mesmo tempo em que recoloca a sociabilidade e a democracia nos ambientes *online* (Santos, 2020).

O crescimento exponencial dos dados agregado aos resultados positivos da implementação do modelo empírico de aprendizagem da máquina, a técnica de *deep learning*, gerou uma capacidade de prever eventos futuros com uma firmeza inusitada. Reconhecendo-se a capacidade preditiva como um dos fundamentos de qualquer atividade económica e social, essa técnica de IA está a disseminar-se rapidamente e está a ser utilizada no desempenho das mais diversas funções, desde diagnósticos médicos até personalização da informação. Muitos setores de atividade já a adotaram, no mínimo, como parceiro da inteligência especializada de um profissional humano. O debate está centrado nas consequências da *dataficação* para a privacidade, o controlo, a vigilância, mormente para o abalo das instituições modernas da sociedade e, especificamente, para a crise das democracias ocidentais (Harari, 2024).

No princípio de 2018, rebentou o escândalo da Cambridge Analytica que fez despertar decisivamente para a questão complexa do uso e partilha de dados, com óbvias implicações na privacidade dos cidadãos. Fundada em 2013 e tendo estado em parte nas mãos da família de Robert Mercer (com interesses no domínio da IA), a Cambridge Analytica esteve envolvida em quarenta e quatro disputas políticas norte-americanas em 2014. A empresa, por via da colaboração com Ste-



ve Bannon, trabalhou com Donald Trump na corrida empresarial de 2016 nos EUA, bem como com o movimento Leave.EU, uma das organizações que fez campanha pelo Brexit no mesmo ano.

Alexander Nix, que liderava a Cambridge Analytica quando irrompeu o escândalo, defendeu que a reunião de todos esses dados e as estratégias de *microtargeting* beneficiam os eleitores, visto que irão receber mensagens sobre temas que lhes são importantes. Contudo, a matéria não tem sido pacífica, o que motivou alguns grupos de proteção dos direitos digitais a levantar questões relativamente à recolha, armazenamento e partilha de informação privada, com os cidadãos à margem de todo este processo, não lhes dando qualquer capacidade de controlo. Se a imagem do Facebook saiu fragilizada neste caso, a Cambridge Analytica comunicou a insolvência em maio de 2018, com alguns dos seus colaboradores a transitarem, nomeadamente, para a Emerdata, uma empresa do mesmo ramo de atividade. Recorde-se que já em 2017 a revista *Time* informava que o Congresso norte-americano estava a investigar uma suposta colaboração da Cambridge Analytica com os russos, numa tentativa de interferir na campanha presidencial que resultou na vitória de Donald Trump. O relatório insinua que a Cambridge Analytica poderia ter tido um papel central na disseminação da propaganda russa através das suas capacidades de *microtargeting*. De novo, as *fake news*. Não se tratando da única empresa do ramo com clientes especificamente da área política, o escândalo deu-lhe uma enorme visibilidade. Em 2019, antigos responsáveis da Cambridge Analytica admitiram que a mesma trabalhou em mais de 200 eleições no mundo inteiro, concretamente em África (Nigéria), na Europa (Chéquia) e América (Argentina).

Num mundo onde a centralidade dos dados é motivo de disrupção e ao reconhecer-se a influência atual do conceito de dataísmo e em que o fluxo de informações exige uma nova ética social, faz sentido evocar Foucault e o termo biopolítica. O termo contemplaria toda a gama de tecnologias de poder com foco nas populações como conjunto de seres vivos coexistentes e os fenómenos naturais a ela subjacentes. A biopolítica propõe-se a regular e intervir sobre as populações massificando comportamentos, quer quando controla taxas de natalidade ou fluxos de migração,

quer quando se atua no controlo de epidemias e pandemias e contribui com programas para a longevidade. Deste modo, o dataísmo é colocado em posição de destaque, suscitando uma recuperação antes do tempo das ideias de Foucault (Lorenzetto & Teixeira Filho, 2022).

A obra de Foucault propunha-se a analisar a relação entre poder e conhecimento, indicando os institutos de dominação e a forma como se dá o controlo social das instituições sociais - a presença de um elevado fluxo de dados (conhecimento) não visto no seu período de vida leva à conclusão de que mais conhecimento geraria mais poder. A contemporaneidade modifica quem são os efetivos detentores desses dados (conhecimento) e qual o poder exercido sobre o indivíduo (trabalhador, cidadão, consumidor). Numa adaptação aos conceitos de biopolítica e biopoder, enunciados por Foucault, encontramos-nos numa fase de superação do indivíduo enquanto figura sociopolítica e económica construída para uma de qualificação do indivíduo enquanto consumidor (Lorenzetto & Teixeira Filho, 2022: 92-94).

A assimilação desses dados alteraram os senhores do biopoder, deixando o Estado-nação e sua biopolítica de concentrar todo o poder, o qual passa a ser fragmentado em diversas fontes dentre as quais pontificam as grandes empresas transnacionais e, sobretudo, as *big techs*. Estas enormes transformações digitais exigem, outrossim, uma nova resposta do direito. O transconstitucionalismo desafia novas ideias dada a alteração dos portadores do biopoder e a formação capilarizada de biopolíticas, demandando maior cooperação e integração, até mesmo digital, para a responsabilização dos envolvidos em atos ilícitos digitais, tendo sempre o compromisso com os direitos fundamentais no centro de qualquer análise (Lorenzetto & Teixeira Filho, 2022: 96-98). Giuliano da Empoli reconhece a atualidade de Foucault quando escreve que “a multidão, a massa compacta, foi abolida em prol de uma reunião de indivíduos separados, cada um dos quais pode ser seguido nos mínimos detalhes” (Empoli, 2023, p. 153). O algoritmo impõe-se e traça um perfil desenhado a partir da sua vida *online*. As gigantes plataformas vendem “produtos preditivos” às grandes empresas ou às forças políticas. É isto que leva Empoli a falar da política quântica, na qual uma versão do mundo que cada um de nós vê é, com

efeito, invisível aos olhos dos outros, uma vez que a “realidade objetiva não existe”, pois “cada observador determina a própria realidade” (Empoli, 2023, pp. 170-171), haja ou não esse “Grande Outro” que nos guie e condicione (Zuboff, 2020).

É nesta relação direta do utilizador com o mundo através da tecnologia digital, dentro da qual se processam novas relações de poder e onde o *populismo digital* explora politicamente o isolamento e os instintos mais básicos, despoletando narrativas ao serviço da conquista do poder formal, que deixa de fazer sentido a *intermediação*, a que se junta a crise da ideia de *representação*, assistindo-se então a um curto-circuito disruptivo entre instituições e *media*. Nesta era de narcisismo de massas, a democracia representativa, ancorada na intermediação, encontra-se em risco e há uma via aberta para as novas tecnologias jogarem a cartada da desintermediação em todos os domínios, num admirável mundo novo sem filtros (Empoli, 2023: 163).

Os desafios suscitados pela emergência da *big data* ainda são difíceis de avaliar em toda a sua amplitude. De uma coisa temos a certeza: o tema é não só complexo como levanta sérias questões sociais e políticas, às quais o poder político tem de dar respostas. ●

Referências

- Bezerra, A. C. (2024). A ideologia dos dados: algoritmos e inteligência artificial no capitalismo da era digital. *Revista Eptic*, v. 26, n. 2, Mai-Ago, 125-139.
- Comissão Europeia (2020). Uma estratégia europeia para os dados. Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões. Bruxelas, 19.2.2020. COM(2020) 66 final.
- De Mauro, A.; Greco, M.; Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *API Conference Proceedings* 1644, 97-104.
- Empoli, G. da (2023). *Os Engenheiros do Caos*. Lisboa: Gradiva.
- Harari, Y. N. (2017). *Homo Deus*. História Breve do Amanhã. 16ª edição. Lisboa: Elsinore.
- Harari, Y. N. (2024). *Nexus*. História Breve das Redes de Informação. Da Idade da Pedra à Inteligência Artificial. Lisboa: Elsinore.
- Lorenzetto, B. M.; Teixeira Filho, A. C. (2022). Dataísmo e biopoder: dados no centro das decisões. *Gavagai*, Erechim, v. 9, n. 2, Jul/Dez, 90-112.
- Santaella, L. & Kaufman, D. (2021). Os dados estão nos engolindo? *Civitas* 21 (2), maio-agosto., 214-223.
- Santos, J. de A. (Org.) (2020). *Política e Democracia na Era Digital*. Lisboa: Parsifal.
- Ularu, E. G., Puican, F. C., Apostu, A. & Velicanu, M. (2012). Perspectives on Big Data and Big Data Analytics. *Database Systems Journal*, vol. III, nº 4.
- Van Dijk, J. (2014). Dataficação, dataísmo e dataveillance: Big Data entre o paradigma científico e a ideologia. *Vigilância & Sociedade* 12 (2), 197-208.
- Weiner, M.; Kochhar, S. (2016). Irreversible: The Public Relations Big Data Revolution. Institute for Public Relations. https://instituteforpr.org/wpcontent/uploads/IPR_PR-Big-Data-Revolution_3-29.pdf.
- Zuboff, S. (2020). *A Era do Capitalismo da Vigilância*. A Disputa por um Futuro Humano na Nova Fronteira do Poder. Lisboa: Relógio D'Água.