

Laura Neiva & Helena Machado

lauracnoo@gmail.com; hmachado@ics.uminho.pt

CENTRO DE ESTUDOS DE COMUNICAÇÃO E SOCIEDADE  
(CECS), UNIVERSIDADE DO MINHO, PORTUGAL

## ***Big Data* – PARADIGMA TECNOLÓGICO QUE FACILITA A PRODUÇÃO DE (DES)INFORMAÇÃO? ASPETOS ÉTICOS E OPERACIONAIS**

### **RESUMO**

Assiste-se hoje à crescente visibilidade do designado *Big Data*. Trata-se de uma técnica que agrega conjuntos massivos de dados, manuseando um volume assinalável de informação, que é analisado a uma velocidade sem precedentes e em tempo real. Enquanto fenómeno social, apresenta três vertentes: a interação entre tecnologia (maximização do poder computacional e precisão algorítmica), as potencialidades de análise (identificação de padrões) e os valores sociais e culturais que rodeiam o *Big Data* (nomeadamente, a crença generalizada que grandes conjuntos de dados oferecem fontes de conhecimento e de previsão mais objetivas e fundamentadas).

*Big Data* pode ser uma fonte de racionalização de acontecimentos, com potencial para aumentar a eficiência e melhorar a precisão da previsão em várias áreas da vida social. No entanto, pode aprofundar desigualdades sociais e económicas pré-existentes e apresentar desafios diversos à privacidade e outros direitos fundamentais. Além disso, a conjugação massiva de dados heterogêneos pode levar a conclusões erradas. A quantidade de dados recolhidos não corresponde à representatividade dos mesmos, mas antes à abrangência da cobertura sobre a vida contemporânea dos indivíduos. Assim, pode também potenciar a produção de desinformação sobre os titulares dos dados recolhidos, por conduzir a conclusões erradas.

Procurando contribuir para este campo de análise, o presente texto apresenta um estudo empírico realizado junto de diferentes profissionais em 25 países da União Europeia, que visou mapear e compreender as representações sociais acerca da potencial aplicação de *Big Data* no campo da investigação criminal. O objetivo central foi aceder às suas expectativas em relação ao potencial impacto das tecnologias na luta contra a criminalidade transfronteiriça. Nestas representações sociais, os diferentes entrevistados tecem considerações éticas acerca do fenómeno *Big Data*, contribuindo para o debate contemporâneo em torno dos direitos, liberdades e garantias.

### **PALAVRAS-CHAVE**

*Big Data*; ética; desigualdades; desinformação

## 1. INTRODUÇÃO

O *Big Data* é um fenómeno que envolve recolher, armazenar, partilhar, avaliar e atuar sobre informações criadas e distribuídas por tecnologias e redes baseadas em sistemas informáticos (Gandy Jr, 1989; Herschel & Miori, 2017; Lyon, 2014; Wood & Ball, 2006). Em fevereiro de 2012, o *The New York Times* anunciava a chegada da era do *Big Data* – *The Age of Big Data* (Lohr, 2012); no ano seguinte, o *The Guardian* identificava a técnica como a *revelação por números* (Burmester, 2013); e em 2017, o *The Economist* noticiava o *Big Data* como *uma nova mina de ouro* (The world's most valuable resource, 2017). Consequentemente, muitos se debruçam sobre esta possibilidade de acesso a enormes quantidades de informações produzidas por e sobre pessoas, coisas e interações. Subsequentemente, surgem debates em torno dos potenciais benefícios e riscos da análise de dados diversos, que podem ir desde sequências genéticas, a interações entre indivíduos e grupos, passando por registos de saúde, registos telefónicos e traços digitais deixados pelas pessoas (boyd & Crawford, 2012).

Tratando-se eminentemente de um *mundo digital*, o *Big Data* é frequentemente apresentado como o *mundo dos dispositivos de geração de dados online*. Trata-se de um mundo concebido como profundamente marcado pela fluidez, mobilidade e aceleração contínua, convertidos em processos de rotina (Brayne, 2017; Frade, 2016; Youtie, Porter & Huang, 2016). Atualmente, as sociedades evoluem paralelamente a um processo de globalização que requer um alto nível de controlo sobre o espaço físico e tempo, resultando na aceleração do ritmo da vida e das temporalidades organizacionais – e isto tem consequências. A *datificação* – a conversão de toda a informação em dados analíticos – tornou-se um paradigma aceite como forma de compreender o comportamento social e a sociedade. Com o advento dos sites, das redes sociais e com o aperfeiçoamento de todo o arsenal tecnológico do mundo digital, expande-se uma indústria de dados que os converte em algo precioso, procurando aumentar a eficácia e eficiência da previsão em várias áreas da vida social (Van Dijck, 2014).

Atualmente, circulam diferentes visões sobre a temática do *Big Data*, maioritariamente difundidas pelos meios de comunicação social – importantes instrumentos de construção de sentidos e significações em torno dos mais diversos assuntos. Consequentemente, urge debater as diferentes perspetivas, temporalidades, especialidades e materialidades que representam os bancos de dados, os seus objetivos e finalidades. Com base neste enquadramento, o presente texto visa explorar as representações

sociais sobre as questões éticas e operacionais desta técnica potencialmente aplicada à investigação criminal transnacional.

## 2. BIG DATA ENQUANTO POTENCIAL TÉCNICA DE INVESTIGAÇÃO CRIMINAL

A definição de *Big Data* não é consensual, mas a maior parte dos autores utilizam o conceito para se referirem aos conjuntos de dados que são recolhidos, analisados, convertidos em algoritmos, categorizados numericamente e identificados por via de um índice calculado para, posteriormente, extrair informação destes dados que informe e oriente as políticas criminais (Chan & Moses, 2017; Drewer & Miladinova, 2017; Gonçalves, 2017; Lefèvre, 2017; Lyon, 2014; Matzner, 2016; Wood & Ball, 2006). As fontes destes dados podem surgir de diversas formas, entre elas: 1) dirigidos (por via dos circuitos de câmaras de vigilância, por exemplo); 2) automatizados (localização geográfica detetada pelo telemóvel, por exemplo); e 3) voluntários (por exemplo, a cedência de dados pessoais por parte dos indivíduos) (Chan & Moses, 2015; Lupton & Michael, 2017; Lyon, 2014).

Na União Europeia (UE) ainda não se consagrou enquanto técnica de investigação criminal, não encontrando aplicação prática. Em contrapartida, nos Estados Unidos da América (EUA) encontra-se amplamente desenvolvida. Os grandes dados ao serviço da prevenção e combate à criminalidade materializam-se em estratégias de: 1) “análise de rede” onde os agentes policiais têm conotados, no seu mapa digital, todos os indivíduos pontuados com alto risco e, depois, analisam as ligações que esses indivíduos têm a lugares e pessoas; 2) uso crescente de pontuações de risco que cotam indivíduos; 3) atividade policial mais preventiva; 4) supervisão sistemática de um elevado número de pessoas; 5) acesso a informações pessoais através dos dados e o cruzamento destes; 6) inserção mais facilitada de indivíduos nos bancos de dados da lei; e 7) os sistemas policiais anteriormente separados, fundiram-se, facilitando a disseminação da vigilância entre muitas e diferentes instituições policiais (Brayne, 2017).

Toda esta panóplia tecnológica levanta questões operacionais, desde logo, acerca da fiabilidade dos dados obtidos e das conclusões retiradas das análises desses dados. Também os desafios éticos e de cidadania suscitados pelo *Big Data*, por via de uma vigilância permanente, invisível e penetrada em todos os interstícios sociais (Matzner, 2016). Isto tem consequências na dificuldade em encontrar um equilíbrio entre a proteção dos direitos dos cidadãos e a necessidade coletiva de segurança (Matos, 2018).

Nos últimos anos, intensificaram-se os debates em torno das implicações sociais, económicas, políticas e éticas decorrentes da circulação e partilha de informação em grande escala e da recolha massiva, utilização e partilha de dados pessoais (Matos, 2018). Enquanto o *Big Data* permite que rapidamente se recolham, analisem e explorem informações, também pode permitir o acesso a dados que comprometem a privacidade dos indivíduos, de modo deliberado ou inadvertido (Herschel & Miori, 2017).

### 3. DEBATES ÉTICOS EM TORNO DE *BIG DATA*

Em abril de 2014, o *The New York Times* anunciava os *Oito (não, nove!) problemas com Big Data* (Marcus & Davis, 2014), alegando que se estava a cometer um erro com o seu uso, visto que se verificava uma atmosfera controversa no planeta dos direitos humanos. A jurista Maria Eduarda Gonçalves escrevia, em março de 2018, no jornal Público, uma peça noticiosa referindo que não há dúvidas de que o *Big Data* desafia o direito fundamental à proteção dos dados pessoais e os princípios reconhecidos pela legislação europeia e inclusive pela Carta dos Direitos Fundamentais (Gonçalves, 2018). Na perspetiva da autora, representa um desafio para o respeito pelos princípios do consentimento (os dados pessoais devem ser processados apenas se o titular dos dados tiver dado o seu consentimento prévio e explícito nesse sentido); finalidade (os dados pessoais só devem ser coligidos para fins específicos, explícitos e legítimos e não devem ser processados de modo incompatível com esses fins); e minimização (o processamento dos dados deve restringir-se ao mínimo indispensável). Estes princípios tornam-se difíceis de cumprir devido à automatização inerente à exploração, análise e reutilização de imensos dados (Gonçalves, 2018).

A recolha de dados é, muitas vezes, automática e passiva (Nunan & Di Domenico, 2013, citados em Ball, Di Domenico & Nunan, 2016), criada por sensores automáticos, como por exemplo, telefones. Consequentemente, as inferências sobre o comportamento humano são extraídas da análise desse fluxo de dados. Assim, surgem questões acerca deste funcionamento, visto que esta recolha de dados compromete a privacidade individual e tem implicações nas relações sociais mais amplas (Bartlett, Lewis, Reyes-Galindo & Stephens, 2018; boyd, 2010, citada em Ball et al., 2016; Coll, 2014). No fundo, permite uma extensão da vigilância comercial e governamental em diferentes aspetos da esfera privada (Ball et al., 2016). Aqui levantam-se questões sobre privacidade individual e exposição

à vigilância relativamente ao autónomo fluxo de dados que o indivíduo não pode controlar (Ball et al., 2016; Coll, 2014).

Uma consequência prática do *Big Data* é a sua dependência, cada vez maior, da análise baseada em algoritmos. Isto representa, por um lado, um afastamento relativamente ao segmento intuitivo de estabelecer um perfil abstrativo do indivíduo; e, por outro, uma aproximação a análises preditivas e modelos quantitativos ajustados para prever comportamentos humanos individuais (Ball et al., 2016). Além disso, a presença de algoritmos converte a prática numa realidade eticamente desafiadora, devido à complexidade da análise e da tomada de decisão (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016). De facto, trata-se de um grande volume de dados, no entanto quando nos referimos a um grande número de dados, o que o termo sugere – grande quantidade, em termos numéricos de fontes de informação – não corresponde ao que representa. O que quer referir-se é à abrangência de cobertura sobre a vida contemporânea, a omnipresença, o conhecimento de um registo quase completo de vidas individuais, que remove a necessidade de decisões *a priori* sobre o início da atividade da vigilância (Ball et al., 2016). E ainda no que concerne aos dados, importa descortinar de onde vêm estes dados, como devem ser interpretados, qual a direção a tomar e não os incluir em estudos sem nexos (boyd & Crawford, 2012). Para além disso, enquanto conjuntos de dados grandes que podem ser modelados, os dados, geralmente, são reduzidos ao que pode ser ajustado num modelo matemático. No entanto, fora do contexto, os dados perdem significado e valor. A capacidade de representar as relações entre as pessoas num gráfico não significa que elas transmitam o equivalente (boyd & Crawford, 2012). Os níveis de permissão, em termos de recolha de dados, anteriormente disponíveis apenas em ambientes políticos rigorosamente controlados, estão agora disponíveis universalmente (Ball et al., 2016). No entanto, *Big Data* não se refere apenas a conjuntos de dados grandes, nem a ferramentas e procedimentos usados para manipulá-los e analisá-los, mas também para uma mudança computacional no pensamento e pesquisa (boyd & Crawford, 2012).

Consequentemente, surgem questões: os dados de pesquisa em larga escala ajudarão a criar melhores ferramentas, serviços e bens públicos? Ou será que vão inaugurar uma nova onda de violações de privacidade? Será a análise de dados uma ajuda para entender comunidades online e movimentos políticos? Ou será usado para investigar os manifestantes e suprimir o seu discurso?

Dada a expansão de *Big Data* como um fenómeno social e técnico, é necessário interrogar criticamente os seus pressupostos e preconceitos (boyd & Crawford, 2012; Coll, 2014). Levantam-se questões éticas e operacionais sobre o fluxo de informação, o acesso a uma grande variedade de dados e a enorme vigilância que é sediada sobre o ser humano. E é sempre crucial considerar a perspetiva do observado e do vigiado – de que forma o indivíduo se apercebe e como, de que está a ser alvo de uma observação criteriosa (Ball et al., 2016)?

Além disso, importa ter em conta que o *Big Data* não é autoexplicativo, ou seja, o facto de o software denotar uma correlação não significa que a mesma se verifique. Este facto realça a necessidade da pesquisa das causas dos fenómenos para uma melhor compreensão dos objetos em estudo (Amoore, 2011; boyd & Crawford, 2012; Chan & Moses, 2015; Mittelstadt et al., 2016). Não obstante, grande parte do entusiasmo em torno do *Big Data* provém da perceção de que este oferece acesso fácil a enormes quantidades de dados. Mas quem recebe acesso? Para que fins? Em que contextos? E com que restrições?

Enquanto a pesquisa usa conjuntos de dados provenientes de diversas fontes, como dados das redes sociais e de pesquisas de sites, as fontes desses dados sugerem que o acesso não é direto. Apenas as empresas de redes sociais têm acesso a grandes dimensões de dados pessoais (boyd & Crawford, 2012). Consequentemente, levantam-se questões éticas assinaláveis no que diz respeito à recolha de dados online. O processo de avaliação da ética da pesquisa científica não pode ser ignorado simplesmente porque os dados são aparentemente públicos. Os investigadores devem debruçar-se sobre a ética do processo da recolha de dados, da sua análise e posterior publicação (boyd & Crawford, 2012).

Assim, importa que nos questionemos se um grande número de dados pode revelar a qualidade das informações recolhidas sobre os indivíduos. No âmbito criminal, visto que a aplicação de *Big Data* serve ações preditivas, é necessário refletir acerca do número de ofensores estatísticos que devem ser conotados e analisados para que se desencadeie uma ação policial. Não obstante, é importante estudar o processo de agregação dos dados (Ball et al., 2016; boyd & Crawford, 2012; Lefèvre, 2017; Metcalf & Crawford, 2016). E considerar, também, que o uso de uma grande quantidade de dados pessoais acarreta uma consequência óbvia: a questão da privacidade individual e social e proteção de dados (Coll, 2014; Drewer & Miladinova, 2017).

#### 4. METODOLOGIA

Esta investigação enquadra-se no projeto EXCHANGE, que explora as dimensões sociais, éticas, políticas, culturais e regulatórias do uso de tecnologias de ADN para uso forense na UE. A pesquisa do projeto utiliza uma abordagem multimetodológica de forma a apreender a pluridimensionalidade dos seus objetos de estudo e a privilegiar as perspetivas socialmente construídas pelos vários atores sociais. Esta metodologia inclui: 1) recolha e análise de documentação em várias fontes relevantes, desde legislação e documentação oficial, até artigos científicos, correspondência aberta e websites relacionados com o regime de Prüm; 2) realização de entrevistas; 3) observação em encontros científicos da genética forense e outros de natureza interdisciplinar e outros contextos situacionais relevantes (por exemplo, laboratórios e outros locais de trabalho de entrevistados ligados a partilha transnacional de informação no âmbito de cooperação policial e judiciária); e 4) recolha e análise de casos criminais que envolvem o uso de tecnologias de ADN.

Para o presente artigo, utilizou-se o material recolhido da realização de entrevistas a 140 entrevistados em 25 países diferentes da UE: Alemanha, Áustria, Bélgica, Bulgária, Chipre, Eslováquia, Eslovénia, Espanha, Estónia, Finlândia, França, Holanda, Hungria, Irlanda, Letónia, Lituânia, Luxemburgo, Malta, Noruega, Polónia, Portugal, Reino Unido, República Checa, Roménia e Suécia; realizadas no âmbito do projeto EXCHANGE (2015-2019). Estas entrevistas foram realizadas a profissionais diretamente envolvidos em redes policiais e judiciárias de partilha transnacional de dados genéticos na UE no âmbito da investigação criminal (n= 47), geneticistas forenses (n= 49), membros de organizações não governamentais (n= 11), entidades de supervisão/regulação (n= 11), legisladores (n= 5), profissionais de empresas privadas (n= 2), professores/investigadores (n= 9), investigadores criminais (n= 5) e membros de meios de comunicação social (n= 1).

O objetivo central desta investigação foi analisar as expectativas sociais destes profissionais acerca da potencial aplicabilidade de *Big Data* à investigação criminal transnacional. O tema foi abordado nas entrevistas, perguntando aos participantes: 1) qual a sua opinião relativamente aos usos potenciais que o *Big Data* proporciona; e 2) qual a sua opinião em relação aos potenciais riscos e benefícios associados ao uso de informações biométricas disponíveis e futuras como *Big Data*. Fruto das entrevistas realizadas, as narrativas expressas pelos participantes agregam considerações

amplas acerca deste fenómeno, alargando a lente de análise às questões éticas e operacionais do conceito em estudo.

Esta investigação ancora-se num pendor metodológico essencialmente qualitativo. Seguindo alguns pressupostos da *Grounded Theory* (Strauss & Corbin, 1990), que viabiliza a construção de teorias “fundamentadas” nos próprios dados (Charmaz, 2009). Esta investigação pauta-se por uma articulação e comparação sistemática e paralela entre recolha e análise, dados empíricos e teoria (Strauss & Corbin, 1990). Em concreto, numa primeira etapa, procedeu-se à comparação dos diferentes extratos de entrevistas, com o objetivo de caracterizar as regularidades dos discursos apresentados pelos diferentes entrevistados (Caregnato & Mutti, 2006). Depois, foi feita uma análise categorial temática (Caregnato & Mutti, 2006), sendo que as categorias foram criadas em consonância com os temas que emergem na literatura científica sobre o tema em apreço. Por fim, procedeu-se à categorização dos dados, à classificação dos elementos discursivos de acordo com as suas semelhanças e diferenciações, com reagrupamento e redefinição constante.

Deste processo resultaram as seguintes temáticas categóricas gerais: 1) Informação sobre *Big Data*; 2) Expectativas futuras; 3) Perceção sobre riscos de *Big Data*; e 4) Pareceres éticos e de direitos humanos. No entanto, o presente artigo foca-se na análise parcial de um meta-tema, visando explorar exclusivamente as questões éticas e de direitos humanos mencionadas pelos diferentes entrevistados, no que concerne ao *Big Data*. Na secção seguinte serão apresentados as informações e os extratos das entrevistas realizadas que foram consensualmente considerados pelas autoras como mais ilustrativos de cada uma das categorias que emergiram da análise de conteúdo.

## 5. MAPEANDO E IDENTIFICANDO AS PERCEÇÕES ÉTICAS E DE DIREITOS HUMANOS NO ARSENAL DE *BIG DATA*

*Big Data* é uma técnica em fase de desenvolvimento no âmbito da investigação criminal. Não se encontra, ainda, validada cientificamente para ser implementada e instaurada enquanto meio de investigação criminal, na UE. No entanto, é uma técnica bastante sofisticada em outras áreas de conhecimento, nomeadamente o comércio, os recursos humanos e o marketing. Os entrevistados possuem um discurso consonante com o que a literatura documenta sobre o tema. Os participantes do estudo revelam-se conscientes da ponderação que deve ser feita entre os benefícios que



os grandes dados têm, e as consequências éticas que podem acarretar. Os discursos expressam esta ambivalência entre as vantagens e desvantagens do uso da técnica de *Big Data*, no sentido da validade da informação que obtém, da forma como os dados são recolhidos, analisados e partilhados, bem como, a forma como se alteraram as relações humanas, por via dos dados.

*Big Data* mudou a forma como (re)vemos, agimos e comunicamos (boyd & Crawford, 2012), ao mesmo tempo que tem o poder de informar como entendemos redes e comunidades. Verifica-se uma digitalização das relações humanas e sociais, que é possibilitada pelo tempo contemporâneo digital e pelas comunicações em rede. Com o surgimento, expansão e desenvolvimento das novas tecnologias (Frade, 2016), segundo Aas (2006), a ação humana é codificada, ou seja, convertida em símbolos com significado especial.

A datificação instaura um debate em torno das entidades que têm acesso aos dados, a forma como o acesso se materializa, o modo como a heterogeneidade dos dados que *Big Data* agrega se condensa e que conclusões atinge. Neste sentido, os entrevistados revelam considerações que demonstram a sua preocupação ética em torno destas questões de acesso e cruzamento dos dados:

Para que essas situações [de cruzamento de um grande número de informações] possam passar a ser prática corrente, eventualmente terão de sair mais leis para proteção de determinados interesses dos indivíduos. (...) não pode ser qualquer pessoa que possa ter acesso a esse tipo de informações e ao cruzamento de essas informações, e não pode fazer o cruzamento assim aleatoriamente (...), tem de estar tudo muito bem regulado e muito bem definido.  
(Geneticista Forense)

No que concerne a este acesso aos dados, boyd e Crawford (2012) referem que importa questionar quem tem acesso aos dados, em que contextos, com que restrições e qual a finalidade. O acesso a informações pessoais através dos dados e o cruzamento destes pode comprometer a privacidade dos indivíduos (Herschel & Miori, 2017). Caso este acesso aos dados não seja claramente definido, deixa-se à mercê que seja feito por qualquer pessoa; e quanto maior for o número de pessoas com acesso livre aos dados, maior é a probabilidade de se registarem violações de informação pessoal e lesões aos direitos dos titulares dos dados.

Quanto maior o número de pessoas com acesso a essa informação [*Big Data*], maior será o risco destas informações

serem analisadas ou usadas sem consentimento, para outros fins. (Profissional diretamente envolvido em rede policial e judiciária de partilha transnacional de dados genéticos)

Tendo em conta que se trata de um mecanismo ameaçador do ponto de vista da proteção das liberdades, garantias e direitos, é necessário que a tecnologia seja manuseada com validade científica e consciencialização dos seus perigos e riscos. Os profissionais revelam preocupações e motivos pelos quais esta ferramenta se pode converter numa prática *arriscada e perigosa* do ponto de vista da qualidade e validade dos resultados atingidos (Albert, Laberge & Hodges, 2009; Williams & Johnson 2004).

Acho que há sempre um nível de risco. Porque esses métodos [*Big Data*] não são exatos. (Profissional diretamente envolvido em rede policial e judiciária de partilha transnacional de dados genéticos)

Os entrevistados referem, também, as questões éticas relacionadas com o grande volume de dados que *Big Data* agrega. Este não reflete a qualidade dos dados, sendo necessária a sua validação para que as conclusões sejam fidedignas.

Estou particularmente preocupado/a com o uso nacional de dados biométricos devido à qualidade dos dados. A qualidade dos dados já é um grande problema, com *Big Data* a preocupação é, ainda, maior (...) Eu refiro-me não só à qualidade dos dados, mas também à qualidade dos testes realizados e dos procedimentos que levam aos resultados. (Membro pertencente a Organização Não Governamental/Direitos Humanos)

Esta sensibilidade face aos dados e aos seus procedimentos de recolha são, frequentemente, nomeados pelos entrevistados. Os aspetos referidos pelos participantes concentram-se na natureza heterogénea dos dados e no seu volume, que podem colocar em causa a validade científica dos resultados atingidos. Ou seja, trata-se de uma amostra diversa e numerosa e, a partir desta, os resultados obtidos podem tratar-se de falácias e erros – segundo as seguintes respostas – e não é o pretendido, podendo colocar em risco a técnica e a investigação em causa:

Acho que esse cruzamento [de muitos e diferentes dados] pode ser perigoso. Porque pode-nos induzir em erro, podemos cair em falácias (...) é preciso muito cuidado com esse género de coisas (Geneticista Forense)

Sobre estas questões éticas da recolha e agregação de dados, boyd e Crawford (2012) referem que é crucial que nos questionemos acerca da origem dos dados e de como estes devem ser interpretados. Caso os dados sejam incluídos em estudos sem o adequado contexto, perderão o seu valor e significado *a posteriori* (Ball et al., 2016).

## 6. CONCLUSÃO

Os discursos dos entrevistados tornam evidentes que o desenvolvimento do *Big Data* suscita diversas questões éticas e operacionais relacionadas com o acesso a uma grande variedade de dados e a vigilância massiva a que estão sujeitos os cidadãos. Além disso, denotam, também, a fragilidade de considerar as conclusões obtidas por via das técnicas de *Big Data* como únicas e fiáveis, tendo em conta a possibilidade de se obterem resultados errados. Esta técnica, apesar de agregar um volume considerável de informação heterogénea sobre os cidadãos, não é representativa do ponto de vista da qualidade das fontes destas informações. *Big Data* pode reunir dados tão distintos que as conclusões atingidas podem, facilmente, tornar-se falácias. As respostas dos diferentes profissionais ressaltam as questões que se prendem com o manuseamento cuidadoso que *Big Data* requer, para diminuir as probabilidades de produção de desinformação. Os participantes alertam para o facto desta técnica não ser tomada como uma via única de investigação, mas para ser conjugada com outras técnicas tradicionais de investigação criminal para potenciar a sua eficácia (Neiva, 2019).

As narrativas dos entrevistados enfatizam, também, as questões de acesso aos dados, das entidades responsáveis pelo seu manuseamento, análise e tratamento. Esta conclusão realça a necessidade de balizar, formal e legalmente, o processo pelo qual o acesso aos dados se materializa, bem como, o seu tratamento e análise. Tendo em conta que este arsenal processual tem repercussões nas conclusões obtidas e um procedimento validado cientificamente poderá reduzir potenciais erros (Neiva, 2019).

Analisar este fenómeno crescente – *Big Data* – sob a lente de quem poderá, potencialmente, manusear a técnica é crucial. Nos últimos tempos, o desenvolvimento dos grandes dados, das análises por via destes e das decisões tomadas sob os resultados obtidos tornaram-se mais frequentes. Assim sendo, importa descortinar de que forma é que a técnica se materializa, visto que o seu desenvolvimento e crescente aplicação acarretam consequências para os direitos, liberdades e garantias. O acesso, a maior parte

das vezes, é automático e a produção de informação e análise processam-se a uma velocidade feroz. Tendo em conta que este processo de produção de informação sobre as pessoas pode, potencialmente, orientar políticas de decisão criminal, deve precaver-se a obtenção de conclusões válidas, sob pena de, para além de lesar direitos dos indivíduos, encaminhar ações erradas (Neiva, 2019).

Assim, urge debater este fenómeno, à luz da defesa de princípios de transparência, confiança pública e prestação de contas da parte das agências governamentais. Questionar e problematizar o *Big Data* significa contribuir para construir uma sociedade tecnológica mais democrática e coesa. Importa saber quem tem acesso a estes dados, o que é que estes dados significam, como é que esta análise é distribuída e com que finalidades (Neiva, 2019). Para além disso, importa construir bases sólidas reflexivas no conhecimento acerca da potencialidade do *Big Data* para se constituir exclusivo na produção de informação e na gestão de ações e decisões (boyd & Crawford, 2012). Concluindo, é evidente que *Big Data* representa um desafio para o respeito pelo direito à privacidade. Os grandes dados comprometem regras e deveres morais porque, em muitos aspetos, tornaram-se muito poderosos, penetrantes e também essenciais para a vida quotidiana. Isso cria um desafio moral para sociedades, porque as pessoas querem usar as tecnologias que criam *Big Data*, mas também querem controlar a forma como isso as afeta (Neiva, 2019).

## REFERÊNCIAS

- Aas, K. (2006). The body does not lie”: Identity, risk and trust in technoculture. *Crime, Media, Culture*, 2(2), 143-158. <https://doi.org/10.1177%2F1741659006065401>
- Albert, M., Laberge, S. & Hodges, D. (2009). Boundary-work in the health research field: Biomedical and clinician scientists’ perceptions of social science research. *Minerva*, 47(2), 171-194. <https://doi.org/10.1007/s11024-009-9120-8>
- Amoore, L. (2011). Data derivatives: On the emergence of a security risk calculus for our times. *Theory, Culture & Society*, 28(6), 24-43. <https://doi.org/10.1177%2F0263276411417430>
- Ball, K., Di Domenico, M. & Nunan, D. (2016). Big data surveillance and the body-subject. *Body & Society*, 22(2), 58-81. <https://doi.org/10.1177%2F1357034X15624973>

- Bartlett, A., Lewis, J., Reyes-Galindo, L. & Stephens, N. (2018). The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -omic biology and high-energy physics. *Big Data & Society*, 5(1), 1-15. <https://doi.org/10.1177%2F2053951718768831>
- boyd, d. & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Burmester, C. (2013, 01 de novembro). How behavioural design can overcome the dark side of big data. *The Guardian*. Retirado de <https://www.theguardian.com/sustainable-business/dark-side-big-data-behavioural-design>.
- Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977-1008. <https://doi.org/10.1177%2F0003122417725865>
- Caregnato, A. & Mutti, R. (2006). Pesquisa qualitativa: análise de discurso versus análise de conteúdo. *Texto & Contexto Enfermagem*, 15(4), 679-84. <https://doi.org/10.1590/S0104-07072006000400017>
- Chan, J. & Moses, L. (2015). Is Big Data challenging criminology? *Theoretical Criminology*, 20(1), 21-39. <https://doi.org/10.1177%2F1362480615586614>
- Chan, J. & Moses, L. (2017). Making sense of Big Data for security. *British Journal of Criminology*, 57(2), 299-319. <https://doi.org/10.1093/bjc/azw059>
- Charmaz, K. (2009). *A construção da teoria fundamentada: Guia prático para análise qualitativa*. Porto Alegre: Artmed.
- Coll, S. (2014). Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society*, 17(10), 1250-1263. <https://doi.org/10.1080/1369118X.2014.918636>
- Drewer, D. & Miladinova, V. (2017). The BIG DATA challenge: Impact and opportunity of large quantities of information under the Europol Regulation. *Computer Law & Security Review*, 33(3), 298-308. <https://doi.org/10.1016/j.clsr.2017.03.006>
- Frade, C. (2016). Social theory and the politics of Big Data and method. *Sociology*, 50(5), 863-877. <https://doi.org/10.1177%2F0038038515614186>
- Gandy Jr, H. (1989). The surveillance society: information technology and bureaucratic social control. *Journal of Communication*, 39(3), 61-76. <https://doi.org/10.1111/j.1460-2466.1989.tb01040.x>
- Gonçalves, M. (2017). The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward. *Information & Communications Technology Law*, 26(2), 90-115. <https://doi.org/10.1080/13600834.2017.1295838>

- Gonçalves, M. (2018, 20 de março). Quem controla os nossos dados? *Público*. Retirado de <https://www.publico.pt/2018/03/20/tecnologia/opiniao/quem-controla-os-nossos-dados-1807206>
- Herschel, R. & Miori, V. (2017). Ethics & Big Data. *Technology in Society*, 49, 31-36. <https://doi.org/10.1016/j.techsoc.2017.03.003>
- Lefèvre, T. (2017). Big data in forensic science and medicine. *Journal of Forensic and Legal Medicine*, 57, 1-6. <https://doi.org/10.1016/j.jflm.2017.08.001>
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society* 1(2), 1-13. <https://doi.org/10.1177%2F2053951714541861>
- Lohr, S. (2012, 11 de fevereiro). The Age of Big Data. *The New York Times*. Retirado de <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>
- Lupton, D. & Michael, M. (2017). 'Depends on Who's Got the Data': Public Understandings of Personal Digital Dataveillance. *Surveillance & Society*, 15(2), 254-268. <https://doi.org/10.24908/ss.v15i2.6332>
- Marcus, G. & Davis, E. (2014, 06 de abril). Eight (No, Nine!) Problems With Big Data. *The New York Times*. Retirado de <https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html>
- Matos, S. (2018). Biometria e privacidade: desafios bioéticos na cooperação policial e judicial na União Europeia. In A. Sol & S. Gouveia (Eds.), *Bioética no Século XXI* (pp. 255-286). Charleston: CreateSpace Independent Publishing.
- Matzner, T. (2016). Beyond data as representation: The performativity of Big Data in surveillance. *Surveillance & Society*, 14(2), 197-210. <https://doi.org/10.24908/ss.v14i2.5831>
- Metcalf, J. & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*, 3(1), 1-14. <https://doi.org/10.1177%2F2053951716650211>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21. <https://doi.org/10.1177%2F2053951716679679>
- Neiva, L. (2019). *Big Data na investigação criminal: previsão do risco, vigilância e expectativas sociais na União Europeia*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal. Retirado de <http://hdl.handle.net/1822/61235>

- Strauss, A. & Corbin, J. (1990). *Basis of qualitative research: Grounded theory procedures and techniques for developing grounded theory*. Newbury Park: SAGE Publications.
- The world's most valuable resource is no longer oil, but data. (2017, 06 de maio). *The Economist*. Retirado de <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-201. <https://doi.org/10.24908/ss.v12i2.4776>
- Williams, R. & Johnson, P. (2004). 'Wonderment and dread': Representations of DNA in ethical disputes about forensic DNA databases. *New Genetics and Society*, 23(2), 205–223. <https://doi.org/10.1080/1463677042000237035>
- Wood, D. M. & Ball, K. (Eds). (2006). *A report on the surveillance society*. Retirado de <https://ico.org.uk/media/about-the-ico/documents/1042388/surveillance-society-public-discussion-document-06.pdf>
- Youtie, J., Porter, A. & Huang, Y. (2016). Early social science research about Big Data. *Science and Public Policy*, 44(1), 65-74. <https://doi.org/10.1093/scipol/scw021>

Citação:

Neiva, L. & Machado, H. (2019). *Big Data – paradigma tecnológico que facilita a produção de (des) informação? Aspectos éticos e operacionais*. In S. Pereira (Ed.), *Literacia, Media e Cidadania – Livro de Atas do 5.º congresso* (pp. 229-243). Braga: CECS.