



UnB

LEONARDO CAMBRAIA CORRÊA

ANÁLISE BIOÉTICA DO USO DE BIG DATA EM SAÚDE

BRASÍLIA, 2022



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOÉTICA

LEONARDO CAMBRAIA CORRÊA

ANÁLISE BIOÉTICA DO USO DE BIG DATA EM SAÚDE

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Bioética pelo Programa de Pós-Graduação em Bioética da Faculdade de Ciências da Saúde da Universidade de Brasília.

Orientadora: Prof. Dra. Monique Pyrrho

Brasília

2022

LEONARDO CAMBRAIA CORRÊA

ANÁLISE BIOÉTICA DO USO DE BIG DATA EM SAÚDE

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Bioética pelo Programa de Pós-Graduação em Bioética da Faculdade de Ciências da Saúde da Universidade de Brasília.

Aprovada em 25/01/2022

BANCA EXAMINADORA

Prof. Dra. Monique Pyrrho
PPGBioética / Universidade de Brasília (UnB)
Presidente

Prof. Dr. Volnei Garrafa
PPGBioética / Universidade de Brasília (UnB)
Examinador interno

Prof. Dr. José Manuel Silvero
Dirección General de Investigación Científica y Tecnológica del Rectorado de la
Universidad Nacional de Asunción
Examinador externo

Prof. Dr. Gabriele Cornelli
PPGBioética / Universidade de Brasília (UnB)
Suplente

À Allanna, pelo amor e apoio. Aos meus pais e irmã, por toda a estrutura. Ao professor Volnei Garrafa, por ter aberto as portas da bioética para um grato engenheiro. À minha caríssima orientadora Monique Pyrho, por seu incomparável suporte ao longo desta dissertação.

“Os poderes sem precedentes que a ciência agora torna possíveis devem ser acompanhados por níveis sem precedentes de atenção e interesse éticos por parte da comunidade científica — bem como pela educação pública mais abrangente no que diz respeito à importância da ciência e da democracia.”

Sagan C. O mundo assombrado pelos demônios. São Paulo: Companhia das Letras, 2006, p. 471.

RESUMO

A inédita capacidade de análise e predição das tecnologias de Big Data é bastante promissora no campo da saúde. O uso massivo de dados pessoais, no entanto, traz riscos individuais e coletivos significativos. Esta dissertação tem por objetivo compreender o presente estado da reflexão ética sobre o uso de Big Data em saúde e as principais implicações já identificadas. Trata-se de estudo comparativo entre o Relatório sobre Big Data e Saúde do Comitê Internacional de Bioética (IBC) da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) e a literatura científica nacional, regional e internacional sobre o tema. A análise dos artigos foi feita com apoio da ferramenta IRaMuTeQ. Como resultado, observou-se que o Relatório identifica como os principais aspectos éticos das novas tecnologias: a autonomia, a privacidade e o equilíbrio entre interesses individuais e coletivos. A literatura, por sua vez, é ainda incipiente, mas já aponta dimensões digitais (técnicas), sociais, econômicas e regulatórias com aspectos éticos importantes. Dentre eles figuram: a incerteza e imprevisibilidade, opacidade, justiça e desigualdade, consentimento, modulação do comportamento, validade e apropriação do conhecimento, e privacidade e confidencialidade, entre outros. O Relatório introduz as principais questões éticas levantadas pelo tema, incluindo a justiça e não-discriminação e a questão ambiental, esta última ausente na literatura analisada. O IBC, no entanto, não tratou de temas como as alterações na condição humana, decorrentes da intensificação da socialização digital, e a modulação comportamental dos usuários. Contudo, é a ausência da discussão a respeito da comoditização da privacidade que mais se faz sentir no Relatório. O conceito de comoditização de privacidade proposto por Zuboff evidencia a expropriação e transformação da privacidade em ativo financeiro. Sem controle social e governança democrática, o fenômeno pode se tornar o principal responsável pelo incremento futuro das desigualdades sociais já existentes. A presente dissertação se propõe a contribuir para o desenvolvimento de reflexão crítica e pautada em contextos regionais periféricos, praticamente nula na amostra analisada, por considerar este um dos instrumentos necessários para um uso mais justo dessas novas tecnologias.

Palavras-chave: Big Data; Bioética; Privacidade, IRaMuTeQ.

ABSTRACT

The unprecedented analytical and predictive power of Big Data technologies holds great promise in the field of healthcare. The massive use of personal data, however, brings significant individual and collective risks. This dissertation aims to understand the present state of ethical thinking on the use of Big Data in healthcare and the major implications already identified. This is a comparative study between the Report on Big Data and Health of the International Bioethics Committee (IBC) of the United Nations Educational, Scientific and Cultural Organization (Unesco) and the national, regional and international scientific literature on the subject. The analysis of the articles was done with the support of the IRaMuTeQ tool. As a result, it was observed that the Report identifies as the main ethical aspects of new technologies: autonomy, privacy, and balance between individual and collective interests. The literature, in turn, is still incipient, but already points out digital (technical), social, economic, and regulatory dimensions with important ethical aspects. These include uncertainty and unpredictability, opacity, fairness and inequality, consent, behavior modulation, validity and appropriation of knowledge, and privacy and confidentiality, among others. The Report introduces the main ethical issues raised by the topic, including justice and non-discrimination and the environmental issue, the latter absent in the literature reviewed. The IBC, however, did not address issues such as changes in the human condition, arising from the intensification of digital socialization, and the behavioral modulation of users. However, it is the lack of discussion regarding the commodification of privacy that is the most felt absence in the Report. The concept of commodification of privacy proposed by Zuboff highlights the expropriation and transformation of privacy into a financial asset. Without social control and democratic governance, the phenomenon may become the main responsible for the future increase in existing social inequalities. The present dissertation proposes to contribute to the development of critical and guided reflection in peripheral regional contexts, practically null in the sample analyzed, for considering this one of the necessary instruments for a fairer use of these new technologies.

Keywords: Big Data; Bioethics; Privacy; IRaMuTeQ.

LISTA DE ILUSTRAÇÕES

Figura 1 – Visão geral do processo metodológico.	18
Figura 2 – Processo básico de inserção de dados no IRaMuTeQ.	22
Figura 3 – Etapas da divisão de classes de palavras.	24
Figura 4 – Representação do processo de classificação hierárquica descendente operada pelo IRaMuTeQ.	25
Figura 5 – Quantidade de artigos selecionados por ano.	35
Figura 6 – Quantidade de artigos selecionados por país.	36
Figura 7 – Quantidade de artigos selecionados por área do primeiro autor.	37
Figura 8 – Dendrograma com as classes geradas e suas principais palavras.	38
Figura 9 – Representação num plano cartesiano das palavras de cada classe.	41
Figura 10 – Comparação da tendência entre a produção técnica e a discussão ética sobre o uso de Big Data em saúde.	63
Figura 11 – Gráfico de Zipf.	138
Figura 12 – Formas presentes na Classe 1.	139
Figura 13 – Formas presentes na Classe 2.	140
Figura 14 – Formas presentes na Classe 3.	140
Figura 15 – Formas presentes na Classe 4.	141
Figura 16 – Variáveis identificadas na Classe 1.	142
Figura 17 – Variáveis identificadas na Classe 2.	142
Figura 18 – Variáveis identificadas na Classe 3.	143
Figura 19 – Variáveis identificadas na Classe 4.	143

LISTA DE TABELAS

Tabela 1 – Lista de desafios do Relatório do IBC sobre Big Data e Saúde (4).	29
Tabela 2 – Lista de aspectos separados por classes.	43
Tabela 3 – Correspondência entre os desafios do Relatório do IBC sobre Big Data e Saúde (4), as dimensões atingidas por estes e os respectivos aspectos éticos identificados na literatura.	68

LISTA DE ABREVIATURAS E SIGLAS

AFC	Análise Fatorial de Correspondência
BIREME	Biblioteca Regional de Medicina
DeCS	Descritores em Ciências da Saúde
DUBDH	Declaração Universal sobre Bioética e Direitos Humanos
IBC	<i>International Bioethics Committee</i>
IRaMuTeQ	Interface de R para Análise de Texto Multidimensional e Questionário
LGPD	Lei Geral de Proteção de Dados
LILACS	Literatura Latino-Americana e do Caribe em Ciências da Saúde
MeSH	<i>Medical Subject Headings</i>
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OMS	Organização Mundial da Saúde
ONU	Organização das Nações Unidas
OPAS	Organização Pan-Americana da Saúde
ST	Segmento de Texto
Unesco	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

SUMÁRIO

1	INTRODUÇÃO	1
1.1	BIG DATA E BIOÉTICA: PRIMEIRAS ANÁLISES.....	3
1.2	ESTRUTURA DA DISSERTAÇÃO	5
2	OBJETIVOS	7
2.1	OBJETIVO GERAL	7
2.2	OBJETIVOS ESPECÍFICOS.....	7
3	REVISÃO DE LITERATURA.....	8
3.1	BIG DATA EM SAÚDE	8
3.2	BIOÉTICA	13
4	METODOLOGIA	17
4.1	DETERMINAÇÃO DA BASE COMPARATIVA DE ANÁLISE DE MORALIDADE	18
4.2	BUSCA ATIVA DE ARTIGOS CIENTÍFICOS E DETERMINAÇÃO DA AMOSTRA PARA ANÁLISE	19
4.3	ANÁLISE APOIADA POR SOFTWARE.....	21
4.3.1	Tratamentos dos Textos	21
4.3.2	Análise de Classificação Hierárquica Descendente	23
4.3.3	Análise de Especificidades	25
4.4	IDENTIFICAÇÃO DOS DESAFIOS PARA POSTERIOR ANÁLISE COMPARATIVA.....	26
5	RESULTADOS.....	28
5.1	RELATÓRIO DO IBC SOBRE BIG DATA E SAÚDE.....	28
5.2	CARACTERÍSTICAS DO PROCESSO DE DETERMINAÇÃO DO <i>CORPUS</i> DE ANÁLISE DA LITERATURA CIENTÍFICA	35
5.3	ANÁLISES REALIZADAS COM O APOIO DO SOFTWARE IRAMUTEQ..	37
5.3.1	Análise de Classificação Hierárquica Descendente	37
5.3.1.1	Classe 1 – Dimensões Digitais	38
5.3.1.2	Classe 2 – Dimensões Sociais.....	39
5.3.1.3	Classe 3 – Dimensões Econômicas.....	39
5.3.1.4	Classe 4 – Dimensões Regulatórias	40
5.3.1.5	Análise Fatorial por Correspondência.....	40
5.3.2	Análise de Especificidades	41
5.4	IDENTIFICAÇÃO DOS ASPECTOS ÉTICOS NA LITERATURA CIENTÍFICA RELACIONADOS AO USO DE BIG DATA EM SAÚDE.....	42

5.4.1	Classe 1 – Dimensões Digitais	43
5.4.2	Classe 2 – Dimensões Sociais	48
5.4.3	Classe 3 – Dimensões Econômicas	55
5.4.4	Classe 4 – Dimensões Regulatórias.....	56
6	DISCUSSÃO	61
6.1	CONTEXTUALIZAÇÃO DO ESTADO ATUAL DA LITERATURA	62
6.2	ANÁLISE COMPARATIVA.....	68
6.3	A COMODITIZAÇÃO DA PRIVACIDADE.....	75
7	CONSIDERAÇÕES FINAIS	81
	REFERÊNCIAS	84
	ANEXO A – PROPOSTA DE TRADUÇÃO DO AUTOR DO “RELATÓRIO DO IBC SOBRE BIG DATA E SAÚDE”	93
	ANEXO B – LISTA DOS ARTIGOS SELECIONADOS	134
	ANEXO C – RELATÓRIO DE ANÁLISE DE DADOS DO IRAMUTEQ	138

1 INTRODUÇÃO

Todas as áreas da experiência humana já são atingidas pela utilização de Big Data. O termo é utilizado para designar um novo paradigma na gestão de informação em função da crescente capacidade de processamento de grandes, complexos e conectáveis volumes de dados (1).

Os grandes dados – Big Data – crescem e variam muito rapidamente, e derivam de muitas fontes. Além da diversidade de origem, o termo Big Data também serve para denominar muitos recursos e características de diferentes soluções tecnológicas, de modo que não existe uma definição compreensiva e amplamente aceita sobre o que é Big Data (2).

Entretanto, existe algum consenso em torno dos 5 Vs, como são chamados os 5 atributos que caracterizam as tecnologias expoentes do fenômeno Big Data: Volume, Variedade, Velocidade, Validade e Valor. A presença dessas cinco características pode ser considerada como requisito para que determinada tecnologia de coleta, análise e processamento de dados seja considerada Big Data (3).

O aspecto *volume* refere-se ao fato de que, contemporaneamente, a esmagadora maioria dos dados, mais de 99% deles, encontra-se em formato digital. Desta maneira, as tecnologias precisam ser capazes de tratar um volume massivo de dados (2).

A *variedade* descreve a diversidade de tipos e fontes de dados tratados. A obtenção de dados em Big Data inclui dados do tipo estruturado (aqueles que derivam de coletas com campos fixos e que, por isso, seguem padrões pré-determinados); dados semiestruturados (dados que não foram obtidos a partir de campos fixos, mas que possuem identificadores ou *tags* que diferenciam cada um de seus elementos) e não-estruturados (dados variados obtidos das atividades aleatórias dos usuários, sem seguir padrões) (2).

No campo da saúde, por exemplo, os dados podem ser oriundos de coleta realizada no cuidado individual do paciente, dados laboratoriais, dados genéticos, dados de saúde pública, dados provenientes de seguradoras, dados coletados por pesquisadores, dados de estilo de vida etc. Adicionalmente, dados de todas essas fontes podem ser classificados de diferentes maneiras, de acordo com diferentes

critérios, como dados pessoais, dados anonimizados, metadados, dados primários e secundários (4).

O atributo de *velocidade* descreve a necessidade de que tais dados possam ser tratados em alta velocidade, de modo que soluções tecnológicas atuais possam gerar análises e até mesmo produzir recomendações imediatas (2).

O termo *validade* refere-se ao requisito de qualidade e confiabilidade dos dados, implicando no requisito de adequação e precisão do conteúdo (2).

Por último, o termo *valor* traduz a significância assumida por determinado conjunto de dados para a compreensão de uma determinada questão, em outras palavras, descreve a imprescindibilidade de determinado conjunto de dados para abordar o objeto de interesse (2).

Quanto ao uso da tecnologia, no campo da saúde, é fato que a compilação e a análise de dados oferecem orientação indispensáveis para o planejamento e decisões das práticas, principalmente no campo da saúde pública, numa temporalidade que remonta ao menos desde o Século XVIII. De alguma maneira, também é possível afirmar que implicações éticas relacionadas ao uso de dados pessoais vêm sendo apontadas desde então (5). O papel crucial da organização de sistemas de vigilância sanitária e mecanismos como notificação compulsória de doenças no enfrentamento de fenômenos epidemiológicos, como epidemias e o combate a doenças endêmicas, é amplo consenso. Entretanto, é também conhecido o risco de exposição dos registros, e suas graves consequências de violação de privacidade e estigmatização dos pacientes. Nesse sentido, são vastamente relatados na literatura eventos que envolvem vazamentos de dados provenientes de vigilância sanitária sobre HIV/AIDS e tuberculose, por exemplo (6).

O interesse nesse no fenômeno Big data, contudo, não resulta somente da quantidade de dados processados, mas da potencialidade de entrecruzamento de informações dos mais diversos setores. Testemunha-se uma habilidade inédita de interpretação e predição de comportamentos humanos. Na área da saúde, por exemplo, já existem aplicativos que monitoram dados pessoais e informações corporais, que permitem prever padrões e sugerir hábitos mais saudáveis de forma individualizada aos seus usuários (1).

Dados ambientais, financeiros, geográficos, de mídias sociais e da saúde podem ser compreendidos em interface, oferecendo novas possibilidades de análise e aplicações (7), bem como novos e aumentados riscos.

1.1 BIG DATA E BIOÉTICA: PRIMEIRAS ANÁLISES

Sucintamente, as tecnologias de Big Data surgem em resposta à necessidade de lidar com um conjunto crescente de dados que é ou muito volumoso ou muito desestruturado para ser gerido e analisado por meios informacionais tradicionais. As soluções tecnológicas pretendem lidar com um fluxo massivo e ininterrupto de “cliques” na Internet, conteúdo de redes sociais e conteúdo em formatos de áudio e vídeo (8).

A coleta e manipulação de dados pessoais por parte das grandes corporações atuantes no ramo, no entanto, representa a possibilidade de impactos negativos para os direitos humanos (9).

É justamente quando se pensa sobre esse tipo de risco que a aproximação entre o campo da bioética e dos Direitos Humanos se mostra mais importante. A reconhecida afinidade entre o referencial dos Direitos Humanos, postulado na Declaração Universal dos Direitos Humanos (10), e as possíveis implicações éticas do saber tecnocientífico, objeto de interesse da bioética, culminou em 2005 na formulação da Declaração Universal sobre Bioética e Direitos Humanos (DUBDH) (11).

A DUBDH em seus 28 artigos aborda questões éticas “relacionadas à medicina, às ciências da vida e às tecnologias associadas quando aplicadas aos seres humanos, levando em conta suas dimensões sociais, legais e ambientais” (11). O documento reflete a ampliação do escopo da bioética, por meio da adoção dos princípios e valores relacionados aos direitos humanos como marcos referenciais para as reflexões da disciplina.

A expansão da disciplina em um contexto global, promovido pela DUBDH, é resultante de um movimento crítico a leituras individualistas dos dilemas e questões bioéticas. Impulsionados pelas propostas teóricas latino-americanas, com principal destaque para a Bioética de Intervenção, os debates críticos resultaram em um documento que fortaleceu a bioética como um campo interdisciplinar de saberes relacionados à saúde, mas cada vez mais interessado em refletir os problemas concretos, em suas dimensões social e ambiental (12). A partir dessa perspectiva, os aspectos ligados à coleta, análise e uso massivos de dados, sintetizados no termo “Big Data”, são objetos de uma preocupação bioética diversa.

Certamente questões como a liberdade individual, são fundamentais. Contudo, interessa especialmente nesta dissertação a compreensão das dimensões morais da interseção entre o fenômeno Big Data e a saúde humana compreendida em sua acepção mais ampla, incluídos seus condicionantes sociais e ambientais.

Os potenciais benefícios e malefícios gerados pela utilização massiva de dados referentes a sequências genéticas, registros de saúde e outros rastros digitais deixados por indivíduos já têm sido objeto de análise (13). Porém, até o momento, têm se discutido majoritariamente aspectos éticos que não são propriamente exclusivos dessa tecnologia, como, por exemplo, é o caso da questão da privacidade (14).

A possibilidade de obtenção e aplicação de agregados de dados é praticamente infinita e as transações financeiras relacionadas à coleta, cruzamento, agregação e venda desses dados são parte significativa da atividade econômica global. Esse emergente setor corresponde, grosso modo, à origem de todas as grandes acumulações de riqueza surgidas nesse milênio (15).

Os desdobramentos do fenômeno não são somente econômicos, mas atravessam a organização social como um todo. Fala-se em letramento digital (*digital literacy*), em uma divisão digital (*digital divide*) que separa os pobres em dados (*data poor*) dos ricos em dados (*data rich*). Esses termos descrevem uma distribuição desigual dos efeitos da tecnologia e um grande potencial de incrementar iniquidades já agudas em função de uma concentração de informações e de capacidade de processá-las (16).

É em função desses mecanismos injustos que a preocupação bioética em países periféricos não pode se desenvolver de maneira neutral, sob o risco de contribuir com a ampliação das distorções sociais existentes nesses contextos (17). Dessa maneira, a análise bioética das tecnologias de Big Data e sua apropriação pelo campo da saúde precisa contemplar as dinâmicas sociais e econômicas envolvidas no processo.

É adicional fonte de preocupação o fato de que não há no mundo regulamentação legal específica sobre o uso de Big Data (14). As regras e leis nacionais e internacionais existentes tratam da privacidade e confidencialidade dos dados. Tal abordagem, no entanto, não se mostra suficientemente abrangente, uma vez que já há aplicações econômicas e políticas destas novas habilidades

informacionais, com vasto alcance populacional. Diante disso, o presente debate ético e regulatório sobre o tema se demonstra ainda incipiente.

Dessa maneira, é importante entender o que é Big Data e como tais dados são utilizados para analisar suas implicações éticas. É importante não perder de vista que esses resultados não se limitam ao nível individual, mas podem afetar a humanidade e sua maneira de se organizar socialmente. A manipulação de dados com fins de indução de comportamentos de consumo são uma preocupação, mas o uso dessas soluções tecnológicas com fins políticos para a vigilância e controle dos cidadãos é uma ameaça a liberdades individuais e aos processos democráticos (18).

Com intuito de ampliar a discussão e ressaltar a importância do tema, o Comitê Internacional de Bioética (IBC, da sigla em inglês: *International Bioethics Committee*) da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), formado por um grupo de *experts* de diversas nacionalidades, responsável pela produção de análises pioneiras sobre as principais temáticas relacionadas ao desenvolvimento biotecnocientífico (19), em 2017, tratou do uso de Big Data em saúde.

No Relatório do IBC sobre Big Data e Saúde (4), o IBC destacou alguns aspectos técnicos preliminares importantes para a análise deste novo fenômeno, e apontou aquelas que foram identificadas como as principais questões éticas relacionadas ao uso dessas novas tecnologias. Uma tradução deste relatório para o português é proposta por este autor com intuito de aumentar o acesso ao documento e ampliar a discussão no contexto brasileiro. Esta encontra-se no Anexo A deste texto.

Visando colaborar com a compreensão da moralidade do uso de Big Data no campo da saúde, a presente dissertação destina-se a analisar a atual produção acadêmica sobre bioética e Big Data em saúde. O estudo tem como objeto de análise, além do já mencionado Relatório do IBC sobre Big Data e Saúde (4), artigos científicos sobre a mesma temática.

1.2 ESTRUTURA DA DISSERTAÇÃO

Seguindo-se a este Capítulo 1 de **Introdução**, no capítulo 2, **Objetivos**, são apresentados os objetivos da pesquisa descrita neste trabalho.

No capítulo 3, **Revisão de literatura**, é apresentada a revisão bibliográfica oferecida como uma abordagem introdutória aos temas: bioética, Big Data e as implicações éticas do uso de Big Data em saúde.

No capítulo 4, **Metodologia**, é descrita a abordagem metodológica do trabalho. São expostos os critérios de seleção, análise e comparação dos textos analisados. Também é apresentado o software usado como apoio à análise, o tratamento dado aos textos para este uso e quais dos seus recursos foram utilizados. Ao fim do capítulo, esclarecem-se os procedimentos metodológicos de análise qualitativa e comparativa da pesquisa.

No capítulo 5, **Resultados**, são apresentados detalhes acerca da seleção do *corpus* analisado e os resultados da análise apoiada pelo software. Daí, são apresentados resultados qualitativos acerca da análise comparativa dos aspectos listados no Relatório do IBC sobre Big Data e Saúde (4) e aqueles identificados pela literatura científica sobre as implicações éticas do uso de Big Data em Saúde.

No capítulo 6, **Discussão**, explora-se a comparação entre os aspectos éticos identificados na literatura e aqueles enumerados pelo IBC, apontando a natureza dos aspectos que destoam entre os dois corpos de texto comparados.

No capítulo 7, **Considerações finais**, apontam-se os principais aspectos a serem aprofundados do ponto de vista de uma bioética latino-americana comprometida com a promoção da justiça e socialmente interessada.

2 OBJETIVOS

2.1 OBJETIVO GERAL

- Levantar as principais implicações éticas da utilização de Big Data no campo da saúde humana conforme são identificadas pela atual produção científica sobre o tema.

2.2 OBJETIVOS ESPECÍFICOS

- Categorizar as implicações éticas encontradas de acordo com sua temática;
- Comparar as implicações identificadas no Relatório do IBC sobre Big Data e Saúde (4) com as publicações científicas sobre o mesmo tema;
- Discutir, à luz da bioética, as implicações éticas da utilização de Big Data no campo da saúde humana, de forma a compreender aspectos de interesse relacionados a essas aplicações em contextos periféricos e latino-americanos.

3 REVISÃO DE LITERATURA

A utilização de Big Data em saúde tem aspectos éticos significativos, derivados de fatores técnicos específicos dessas novas tecnologias. Visando melhor compreender a dimensão ética desse novo fenômeno, será apresentado a seguir um apanhado introdutório sobre o tema.

Com o intuito de apresentar os referenciais para a leitura bioética sobre o tema, será apresentada também uma breve revisão sobre a bioética e a Bioética de Intervenção, proposta teórica latino-americana e crítica (17) que analisa os problemas bioéticos a partir de uma preocupação centrada nas distorções sociais e as iniquidades em saúde.

3.1 BIG DATA EM SAÚDE

Nos anos 1970, para lidar com as informações necessárias a suas atividades, as corporações possuíam um computador dedicado exclusivamente ao armazenamento e análise dos dados. Ao passo que o fluxo de dados aumentava, essa solução foi perdendo sua praticidade. Nos anos 1980, foi proposto um sistema paralelo de banco de dados, cuja arquitetura previa que cada máquina tinha seu processador e armazenamento. Esse foi o sistema que perdurou até os anos 1990 quando, com o desenvolvimento da Internet, houve uma explosão na quantidade de dados em trânsito (20).

No começo dos anos 2000, a grande quantidade de dados passou inclusive a ser vista como um problema técnico, dado que os computadores da época não tinham poder de processamento suficiente para acompanhar essa quantidade de dados que surgia. No entanto, a capacidade de processamento também assistiu a um progresso muito rápido. Obedecendo a Lei de Moore (segundo a qual, em termos leigos, sem alteração de custo, a cada dois anos, o poder de processamento de processadores de computador aumenta 100%), o problema da falta de capacidade de processamento perdeu em importância e testemunhou-se um grande incremento na capacidade de análises dos crescentes volumes de dados (21).

É desta época que data uma das possíveis primeiras definições de Big Data. O termo aparece em um relatório produzido em 2001, em que um analista, Doug Laney, definiu desafios e oportunidades trazidos pelo aumento do fluxo de dados apresentando um modelo de 3 Vs: volume, velocidade e variedade. Mesmo que tal modelo não tenha sido usado como definição de Big Data originalmente, diversas companhias usaram os 3 Vs para descrever Big Data por muito tempo (20).

As primeiras organizações a tirar proveito de Big Data foram as corporações com foco em operações *on-line*, sendo que muitas das atuais gigantes da tecnologia (Google, Facebook, etc.) nasceram em torno desta tecnologia. Diferente de companhias que já existiam desde antes da virada do milênio, essas companhias, porque estrearam já em um novo contexto, não tiveram de integrar sua área analítica, mudar suas fontes de dados ou alterar suas infraestruturas de tecnologia. Para essas nascentes corporações, as tecnologias de Big Data foram o padrão desde o momento zero (22).

O fenômeno, no entanto, não foi uma realidade que se ergueu do dia para a noite. Muitas grandes corporações do ramo das tecnologias informacionais já vinham adicionando novas formas de dados em seus sistemas por vários anos, de modo que o que chama atenção sobre Big Data não é necessariamente o seu tamanho, apesar de esta ser uma de suas principais características, mas sua falta de estrutura, ou seja, sua capacidade de aproveitamento de novas fontes de dados não estruturadas, sejam estas texto, voz ou vídeo, por exemplo (22).

Após o grande desenvolvimento das corporações com operações baseadas na Internet, em 2011 foi publicado um relatório de pesquisa intitulado *"Extracting Values from Chaos"*. Nele, foi proposta uma definição de Big Data e foram apresentados os possíveis benefícios do seu uso. Esse foi um evento importante para o generalizado despertar de interesse sobre o tema (20).

Desde então, praticamente todas as grandes companhias de tecnologia atuam em Big Data. Esse movimento foi acompanhado pela academia, tanto por disciplinas da área, como engenharia de software e ciência da computação, quanto por diversas outras que utilizam a tecnologia para alavancar a pesquisa em suas respectivas temáticas (20).

O interesse na nova tecnologia reside na possibilidade de exploração de conhecimento de várias fontes de informações para o processo de tomada de decisão. Em gestão empresarial e governamental, as promessas são várias e incluem aumento

de eficiência operacional; decisões estratégicas informadas; melhora no atendimento a consumidores e cidadãos; identificação e desenvolvimento de novos produtos e serviços; melhora na experiência de atendimento e consumo; identificação de novos mercados; diminuição do tempo necessário à introdução de novos produtos no mercado; facilidade nos ajustes de adequação a regulações, dentre outros (23).

De modo geral, as análises de Big Data são utilizadas para tratar conjuntos de dados que, por serem diversos ou grandes demais, se tornam demandas muito difíceis para métodos tradicionais (23). Um dos principais diferenciais das tecnologias de Big Data é justamente a aumentada capacidade de processar grandes volumes de dados não estruturados, em tempo real (20).

Para além do ganho de eficiência quanto a conjuntos de dados já consolidados e existentes, torna-se possível a utilização de dados em tempo real obtidos a partir de sensores e outros tipos de dispositivos. As aplicações incluem análises de fraudes de identidade em tempo real e avaliações de risco médico de pacientes. Além desse tipo de uso pontual, existe também a possibilidade de constante monitoramento, como, por exemplo, a detecção de mudanças da percepção quanto ao interesse em determinado produto ou conjunto de produtos ou o monitoramento de motores de aeronaves (8).

Deste modo, é válido notar que uma das principais mudanças na utilização de Big Data é quanto à temporalidade da análise. De tal modo, em vez de analisar apenas dados passados, analisar dados em tempo real, de forma contínua, aumenta a capacidade preditiva dos sistemas. No campo da saúde, as aplicações são diversas. Foi reportada, por exemplo, como uma das primeiras aplicações no campo da saúde, no Hospital de Toronto para Crianças Doentes (*Toronto's Hospital for Sick Children*, em tradução livre), a utilização de algoritmos de aprendizado de máquina que se destinam a antever padrões séricos que precedem infecções em bebês prematuros (8).

Para obter tais resultados em capacidade preditiva, é necessário o desenvolvimento técnico de diversas disciplinas, em que métodos multidisciplinares se tornam necessários para acessar esse conhecimento. São utilizadas ferramentas matemáticas e técnicas de análise de dados para criar as aplicações de Big Data. Dentre as ferramentas matemáticas estão os métodos de otimização, que consistem em diferentes algoritmos e estratégias computacionais cujo objetivo é aprimorar o processamento de dados utilizados para resolver problemas quantitativos. Muitas

pesquisas têm sido empreendidas para evoluir esses algoritmos dado que a otimização em tempo real é necessária em muitas aplicações de Big Data (23). Um exemplo desses métodos é a técnica de redução de dados, que auxilia na mineração de textos e recuperação de informações, reduzindo de forma significativa os custos computacionais utilizados ao processar textos (24).

Dentre as técnicas de análise de dados mais importantes destacam-se a mineração de dados e o aprendizado de máquina. A mineração de dados é o conjunto de técnicas usadas para extrair padrões dos dados. O termo mineração abarca diversos métodos e técnicas (23). Um exemplo de técnica é o reconhecimento de padrões, que, em termos simplificados, é a busca por estrutura nos dados em análise (25). Já o aprendizado de máquina procura desenvolver algoritmos que sejam capazes de evoluir ao analisar dados e tomar decisões sem intervenção humana. Sendo assim, aprendizado de máquina é o campo de estudo que permite que computadores desempenhem tarefas para as quais não foram explicitamente programados para realizar. O foco do aprendizado de máquina é conseguir classificar e prever, fazendo isso com base em informações obtidas por meio de dados de treino (26). Sucintamente, algoritmos comuns fazem o que são programados/ordenados para fazer, ao passo que algoritmos de aprendizado de máquina são ensinados a reconhecer determinados padrões em conjuntos de dados controlados e em seguida buscam esses padrões que foram aprendidos nos conjuntos de dados que serão analisados.

Ainda, antes mesmo de abordar as implicações éticas do uso de Big Data, é preciso compreender seus aspectos técnicos. Nesse sentido, nota-se que a coleta, armazenamento e análise de dados apresentam dificuldades. Em relação à coleta, o volume de dados vem crescendo muito com o uso cada vez maior de dispositivos portáteis e sensores. Para que as análises tenham sucesso, é necessário que os dados não apresentem inconsistências ou incompletudes (23). Além disso, com o advento de Big Data, a coleta de dados cresceu de forma exponencial, de modo que algoritmos comuns não são capazes de gerir e processar tais dados dentro de um tempo aceitável (27).

Em relação ao armazenamento, mesmo com seu crescimento constante desde os anos 1980, em alguns campos, como o financeiro e médico, muitos dados, criados e obtidos a significativos custos, são hoje frequentemente deletados simplesmente por falta de espaço. Apesar do grande crescimento de poder computacional para

análises, a tecnologia de armazenamento também precisa evoluir para garantir um bom uso da tecnologia como um todo (23).

Quanto à análise, dada a escala do volume de informações, o desafio envolve a capacidade do processamento se dar em tempo real. A demanda por análises de agregados de dados cada vez maiores, em um tempo cada vez menor, tem estimulado a pesquisa e desenvolvimento de melhores técnicas de computação, tanto na linha de desenvolvimento de hardware, quanto na arquitetura de software (23).

Para além dos desafios técnicos, os usos transformam o estar-no-mundo de pessoas reais e isto impõe desafios éticos.

Na saúde, o uso de Big Data é capaz de tornar os serviços de saúde mais eficientes e ainda auxilia na mudança de paradigma ao mudar os cuidados em saúde para um modelo com foco em prevenção, intervenção precoce e uma boa gestão. Uma das aplicações de Big Data em saúde são análises de conjuntos de dados (estruturados ou não) de diferentes fontes, cruzando os dados, com promessas de percepção de tendências e criações de modelos preditivos (28).

Como principais fontes de dados podemos identificar os registros eletrônicos em saúde, resultado do processo de coleta e armazenamento de dados rotineira nos centros de saúde. Eles representam uma importante fonte de dados estruturados sobre pacientes, exames, diagnósticos e medicações utilizadas. Outra fonte de dados que tem crescente importância são as redes sociais, uma vez que os indivíduos compartilham uma quantidade crescente de informações sobre sua saúde. Além de prontuários e redes sociais, dados entrecruzados a respeito de fatores locais como dados climáticos oferecem mais informações para as possíveis análises de Big Data. Adicionalmente, dados sobre a poluição e temperatura ambientais de um local podem informar os diagnósticos e prognósticos de indivíduos e populações (28).

Ainda que as novas aplicações no campo da saúde sejam promissoras, as tecnologias de Big Data trazem consigo implicações éticas.

Além das mais evidentes preocupações com a privacidade dos indivíduos que possuem seus dados coletados e analisados, outras dimensões éticas precisam ser analisadas. Os elevados requisitos tecnológicos necessários para gerar inteligência e conhecimento a partir de dados analisados trazem consigo um grande desequilíbrio de conhecimento e poder, que já atua no agravamento de desigualdades sociais já existentes. Outro grande perigo está na possibilidade de Big Data sugerir resultados com base em correlações feitas sem um amplo entendimento das diversas dimensões

sociais do fenômeno analisado, o que pode atuar reforçando preconceitos e estereótipos de diversas maneiras (14).

Mesmo que muitas das implicações éticas do uso de Big Data em saúde, como a questão da privacidade e dos desafios em ética e pesquisa, não sejam exatamente novas, a crescente propagação dessas novas tecnologias em todas as dimensões da experiência humana demanda uma atenção específica (14).

Os desafios impostos pelo uso de Big Data no campo da saúde humana serão analisados nesta dissertação sob uma óptica bioética. Partindo de uma análise bioética que pretende refletir uma perspectiva latino-americana, importam não somente as implicações de uma situação emergente em bioética, mas também as dimensões morais específicas resultantes da disseminação do uso dessas novas tecnologias em contextos periféricos e desiguais (17).

3.2 BIOÉTICA

Parte importante dos estudos em bioética atualmente são destinados a refletir as dimensões éticas do uso de novas tecnologias no campo da saúde. Refletir a relação entre a capacidade tecnocientífica e possíveis implicações negativas faz parte do repertório da disciplina desde seu começo. A bioética, como é conhecida hoje, remonta aos primeiros textos do cientista estadunidense Van Rensselaer Potter, datados de 1970 e 1971. A motivação inicial do bioquímico e pesquisador no ramo de oncologia foi sua preocupação com o rápido progresso científico tecnológico nos anos que sucederam a Segunda Guerra Mundial. Esse momento constitui uma inflexão histórica única: a partir de então, com o conhecimento obtido a partir de estudos científicos, por ato humano toda a humanidade poderia ser eliminada (29).

É para garantir a sobrevivência humana que a bioética é proposta por Potter. Para ele, o termo designava um campo de estudos destinado a servir de ponte, uma aliança necessária entre conhecimentos das ciências, biológicas principalmente, com as humanidades, com fins de garantir a continuidade da espécie (30).

Sua preocupação, no entanto, era mais ampla e abrangia todo o meio ambiente. O crescimento da população e o uso dos recursos naturais esgotáveis eram preocupações centrais para o autor (30).

Paralelamente, nos anos 1970, e ainda na década anterior, mesmo já existindo o Código de Nuremberg (31) e a Declaração de Helsinki (32), emergiram diversos escândalos envolvendo a pesquisa com seres humanos. Pesquisas que, sem informar os pacientes, submetiam sujeitos e populações vulneráveis a práticas eticamente reprováveis.

Talvez o caso mais grave, com papel crucial na intensificação das preocupações com a ética em pesquisa, tenha sido o caso Tuskegee (33).

Num contexto de aquecidos movimentos pelos direitos civis à época, pressionado pela opinião pública, o governo estadunidense criou a Comissão Nacional para a Proteção de Sujeitos Humanos de Pesquisa Biomédica e Comportamental para estudar o tema (34).

Após três anos, a comissão publicou o Relatório Belmont, que trouxe três princípios éticos básicos para a condução de pesquisas envolvendo seres humanos: a autonomia, a beneficência e a justiça (35). Com base nesse relatório, dos anos 1970 aos anos 1990, a bioética foi sendo difundida ao redor do mundo seguindo uma lógica anglo-saxônica, na qual a autonomia dos indivíduos ganhou preponderância. Isso teve seu ápice com o advento do principialismo, que é uma corrente estadunidense que se baseia em quatro princípios supostamente universais: autonomia, beneficência, não maleficência e justiça proposta por Beauchamp e Childress, em 1979, no livro "*Principles of Biomedical Ethics*" (36).

A partir dos anos 90, no entanto, a proposta passou a receber críticas. A principal delas, apresentada por Clouser e Gert (37), é que os princípios não compõem ou se baseiam em uma teoria ética única, mas são "coletados" de diferentes teorias. Dessa maneira, faltaria ao principialismo uma teoria capaz de relacionar seus princípios adequadamente, o que gera dificuldades na aplicação dos princípios na resolução de problemas práticos. Por essa razão, do ponto de vista prático, a aplicação dos princípios encontra dificuldade, porque na resolução de casos concretos podem entrar em contradição (38).

Azambuja e Garrafa (38), em uma análise do principialismo e da teoria da moralidade comum como proposta de esqueleto conceitual que uniria os quatro princípios em um corpo comum, afirmam que:

... os princípios do principialismo não passariam de um grande resumo histórico das teorias da "justiça" de John Rawls, do "utilitarismo" de John Stuart Mill, da "autonomia" de Immanuel Kant e da "não maleficência" de Bernard Gert (38, p. 637).

Outra importante crítica ao principialismo, apresentada por Garrafa e Porto (17), é que o principialismo pouco faz para resolver problemas de origem social e de desigualdade econômica presentes em países periféricos.

Com o surgimento das críticas ao principialismo e a ineficiência desse referencial ao lidar com problemas de origem socioeconômica, a bioética entra num período de revisão crítica, o que a obriga a entrar em um processo de ampliação conceitual. Conforme Garrafa (29), com essa ampliação, a bioética passa a englobar não apenas os temas biomédicos e biotecnológicos e passa a englobar temas de natureza social, sanitária e ambiental.

O ápice da ampliação conceitual ocorreu com a publicação da DUBDH, em 2005. No texto da DUBDH, ficou clara a impossibilidade do uso de princípios éticos universais, dados as diferentes visões morais e de contexto em diferentes locais. Com isso, verifica-se a existência de várias bioéticas, não uma única bioética globalizada (29).

Nesse contexto, surge a Bioética de Intervenção, uma proposta teórica que busca legitimar uma perspectiva ampla que envolve os aspectos sociais, assim contribuindo na resolução de problemas em países periféricos, ou seja, países que se distanciam do norte geopolítico. A bioética latino-americana nasce territorializada e comprometida com a equidade como mecanismo de combate às desigualdades sociais, de modo que se preocupa de forma constante com as populações vulneráveis (39).

Dado o contexto amplo em que se insere, a Bioética de Intervenção traz uma definição prática acerca dos desafios que busca trabalhar. Além de abordar desafios presentes em países desfavorecidos economicamente, como temas relacionados ao começo e fim da vida, a bioética, sob esse amplo novo olhar, pode ser classificada em duas áreas: a bioética das situações emergentes e a bioética das situações persistentes, conforme proposto por Garrafa (40). Em artigo de 1996, Garrafa (40) expressa pela primeira vez que se tem registro, informação corroborada em relatos orais do próprio autor, a utilização destes termos. A diferença entre as duas se dá por meio de um marco temporal, de modo que os problemas historicamente persistentes ao longo da história da humanidade e que ainda se mostram presentes apesar dos diversos avanços tecnológicos são problemas persistentes e os problemas que surgiram nas últimas décadas, oriundos destes mesmos avanços tecnológicos, são problemas emergentes.

Além da categorização entre situações emergentes e persistentes, a diferença entre os diferentes cenários geopolíticos faz a bioética variar conforme o contexto em que se encontra. São observáveis as diferenças entre os países centrais, do norte geopolítico, e os países periféricos, do sul geopolítico quanto a sua produção bioética.

Na corrente estadunidense principialista, tem ênfase a individualidade, de modo que a bioética é entendida como uma necessidade de regular as relações entre dois sujeitos postos em contato devido a questões vinculadas à saúde. Nesta perspectiva, a saúde figura como uma mercadoria/bem. Já a bioética de matriz europeia, por exemplo, possui uma visão diferente acerca da saúde, reconhecendo esta como um direito individual, ligada à dignidade humana. Por sua vez, a bioética latino-americana considera a saúde um direito humano e em suas análises leva em consideração os processos históricos e as relações de poder na determinação das iniquidades de acesso (41).

É aqui que se revela a importância da compreensão da emergência de novas tecnologias a partir de uma visão latino-americana. A Bioética de Intervenção possui uma capacidade singular de análise de desafios de acordo com a realidade local. Em relação às tecnologias de Big Data, por exemplo, em contextos periféricos, mesmo sem poder usufruir de benefícios diretos, uma parcela vulnerável considerável fornece insumos para a manutenção de mecanismos político-econômicos baseados nessa tecnologia. Por conta disso, analisar Big Data sob uma perspectiva bioética autóctone permite observar não apenas os desafios já delineados na literatura, mas ir além destes, observando a influência de Big Data em um contexto tecnologicamente desenvolvido, mas altamente desigual do ponto de vista social.

4 METODOLOGIA

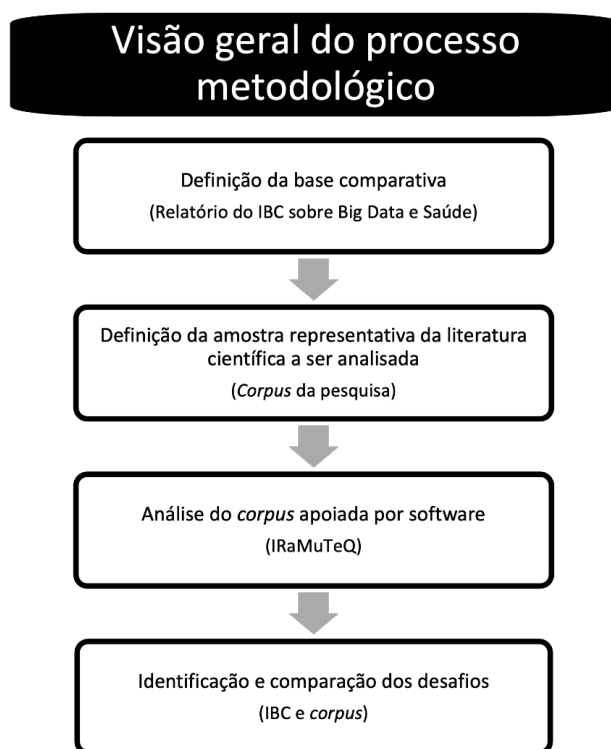
A presente pesquisa consiste em uma análise da atual reflexão ética acerca do uso de Big Data em saúde.

Dessa maneira, o estudo aqui empreendido tem como objetivo dar um passo anterior a uma abordagem prescritiva ao tema. O intuito desta dissertação é oferecer um panorama do presente estado da discussão bioética sobre Big Data, identificando as principais questões já abordadas e apontando para temas que precisam de maior aprofundamento no debate, principalmente em contextos periféricos e desiguais.

Para realizar a análise da moralidade do uso de Big Data em saúde, a presente pesquisa consiste em um estudo comparativo sobre as implicações éticas do uso de Big Data em saúde.

A descrição da abordagem metodológica do trabalho, detalhada a seguir, trata da [1] definição da base comparativa de análise, [2] dos critérios para a busca ativa pelos artigos científicos presentes na comparação, incluindo a determinação da amostra a ser analisada, [3] do processo de análise apoiada por software, incluindo o tratamento prévio dado aos textos e os recursos utilizados neste, e [4] da identificação dos desafios para posterior análise comparativa. Uma visão geral do processo metodológico presente na dissertação pode ser vista na Figura 1.

Figura 1 – Visão geral do processo metodológico.



Fonte: Esquematização realizada pelo autor.

4.1 DETERMINAÇÃO DA BASE COMPARATIVA DE ANÁLISE DE MORALIDADE

Percorrendo o trajeto histórico das reflexões quanto ao uso de Big Data em saúde, o pioneiro Relatório do IBC sobre Big Data e Saúde (4), publicado em 2017, foi definido como uma das bases comparativas da pesquisa.

O IBC (Comitê Internacional de Bioética) da Unesco, criado em 1993, é um comitê acessório formado por 36 experts independentes. O comitê atua para que a promoção do progresso científico configure meio para geração e proteção de direitos humanos, como o respeito à dignidade humana e à liberdade (19).

As sessões do IBC ocorrem ao menos uma vez ao ano e por meio de sessões e grupos de trabalho, o IBC gera recomendações em assuntos específicos, adotados por consenso. O comitê possui dentre seus objetivos a promoção da reflexão sobre questões éticas, o incentivo ao intercâmbio de ideias, o incentivo a ações de conscientização pública, a cooperação com organizações, e a contribuição para a

disseminação dos princípios definidos nas declarações da Unesco sobre bioética. Em sua composição, O IBC leva em consideração a diversidade cultural e representação geográfica balanceada (19).

O comitê em suas discussões pauta as dimensões éticas de temáticas do desenvolvimento tecnocientífico. Seus relatórios apontam para as questões fundamentais de uma determinada temática, constituindo muitas vezes roteiros iniciais, guias preliminares para a reflexão e abordagem de objetos tecnocientíficos ainda pouco estudados do ponto de vista bioético.

O Relatório do IBC sobre Big Data e Saúde (4) possui como objetivo apresentar recomendações para que se atinja toda a potencialidade tecnológica do fenômeno Big Data, salvaguardando o respeito à dignidade humana e direitos e liberdades fundamentais, conforme estabelece o Artigo 3 da DUBDH.

A comparação do documento (4) à literatura científica em bioética sobre o tema tem como finalidade analisar e compreender o presente estado da discussão, identificando eventuais convergências e divergências entre as duas bases comparativas.

4.2 BUSCA ATIVA DE ARTIGOS CIENTÍFICOS E DETERMINAÇÃO DA AMOSTRA PARA ANÁLISE

O levantamento bibliográfico incluiu busca ativa de artigos científicos em contexto nacional, regional (latino-americano) e internacional.

No nível nacional e regional, a pesquisa foi realizada na base virtual “Literatura Latino-Americana e do Caribe em Ciências da Saúde” (LILACS), que é uma base de dados de saúde latino-americana mantida pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, também conhecido por seu nome original Biblioteca Regional de Medicina (BIREME), no endereço <https://lilacs.bvsalud.org>. É válido reforçar que a LILACS engloba a base SciELO, uma base nacional, de modo que a busca nessa base permite uma avaliação tanto nacional quanto regional (42).

No nível internacional, foi utilizada para pesquisa a base de pesquisa PubMed, que é uma das mais abrangentes bases de dados médica internacional, mantida pelo

governo estadunidense, por meio da *National Library of Medicine*, no endereço <https://pubmed.ncbi.nlm.nih.gov> (42).

O método do levantamento bibliográfico em ambas as bases é baseado no uso de descritores. Este consiste em uma convenção classificatória hierárquica da terminologia científica, organizada de forma a otimizar a pesquisa de termos (43). A PubMed utiliza a terminologia “*Medical Subject Headings*” (MeSH); e a LILACS, os “Descritores em Ciências da Saúde” (DeCS).

Os descritores usados como parâmetro para todas as buscas foram: “Big Data”, “Bioética” e “Ética”, em inglês, espanhol e português. Na PubMed, seguindo as opções disponibilizadas pela base, os descritores “*Big Data*” e “*Bioethics*” foram buscados nos títulos e resumos dos artigos; e os descritores “*Big Data*” e “*Ethics*” foram buscados como tópico (*MeSH Terms*) e como subtópico (*Mesh Subheadings*), respectivamente. Na LILACS, a busca dos descritores “Big Data” e “Bioética” se deu nos campos “Assunto de Revista” e “Palavras”.

Foram incluídos os artigos resultantes das especificações de busca acima descritas disponíveis nas buscas até 17 de novembro de 2020.

Foram excluídos artigos publicados em outro idioma além de português, espanhol e inglês. Foram também excluídos os textos que, após leitura exploratória, tenham fugido ao escopo temático do estudo.

Os artigos foram obtidos digitalmente por meio do convênio oferecido pelo Portal de Periódicos CAPES/MEC, mantido pelo governo federal brasileiro. Todos os artigos foram agrupados em uma mesma pasta digital e sofreram tratamento necessário para viabilizar o processamento de seus textos por parte do software utilizado como ferramenta acessória de análise textual.

Ainda, ao realizar a análise dos artigos encontrados, o ano, país de publicação e a área dos primeiros autores destes artigos foram variáveis utilizadas para auxiliar a obtenção de conhecimento sobre o atual estado da produção científica sobre o tema. Isso foi feito por meio da comparação quantitativa do número de artigos por ano e oriundos de cada país e área.

Para determinar a área dos primeiros autores de cada um destes artigos foram analisadas as informações de filiação presente em cada um dos artigos. Nota-se ainda que em alguns casos, devido a falta de informações sobre a filiação dos autores, não foi possível identificar sua área.

4.3 ANÁLISE APOIADA POR SOFTWARE

A análise se valeu de ferramenta auxiliar de análise textual, o software livre IRaMuTeQ. IRaMuTeQ é uma sigla para **I**nterface de **R** para **A**nálise de **T**exto **M**ultidimensional e **Q**uestionário (tradução livre) – na qual **R** é o nome do software estatístico por trás da interface.

O uso do software teve como objetivo apoiar e reforçar a análise por meio de seus recursos, detalhados em seção posterior. O software foi utilizado para analisar a literatura científica em bioética sobre Big Data. Os métodos de busca, critérios de inclusão e exclusão foram detalhados na seção 3.2. O *corpus* definido pelos artigos científicos selecionados recebeu tratamento para o processamento pela ferramenta, conforme detalhado na seção seguinte. Os resultados oferecidos pelo software são então comparados com os desafios relacionados ao uso de Big Data em saúde, conforme listados no Relatório do IBC sobre Big Data e Saúde (4).

É nesse sentido que se diz que esse tipo de estudo consiste em uma análise “apoiada” por ferramenta. Isto porque, embora o software seja capaz de identificar e agrupar semelhanças nos textos, é o pesquisador que lê e interpreta o processo de categorização executado pelo instrumento. Desse modo, por exemplo, para identificar a temática comum de cada classe, são lidos trechos do texto para a posterior denominação do tema daquele conjunto.

A vantagem do uso de uma ferramenta de análise de dados como o IRaMuTeQ é o ganho de escala dentro dos limites de tempo pré-estabelecidos para a pesquisa, o que permite análises comparativas em uma maior quantidade de textos. Outra vantagem é o tratamento estatístico do sistema classificatório, de modo a quantificar a presença de determinadas ocorrências ou a falta delas e testar as similitudes, o que reforça os resultados apresentados (44).

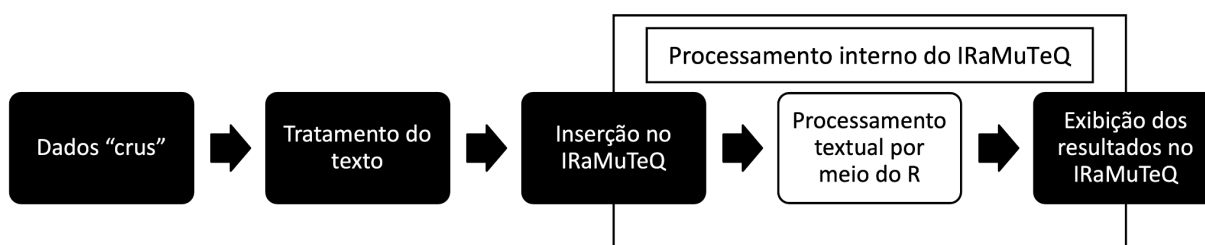
4.3.1 Tratamentos dos Textos

O IRaMuTeQ é um software e atua como uma interface que tem por função receber os dados, realizar o processamento textual por meio do software R (conforme

o algoritmo a ser detalhado na seção relativa às análises) e exibir os resultados de forma compreensível para seu usuário (45). Apesar da interface amigável para a inserção dos dados, estes precisam de um tratamento prévio anterior à utilização e processamento.

Todo o texto analisado foi salvo em formato texto (.txt) e codificação UTF-8. Além disso, todos os caracteres proibidos pelo IRaMuTeQ (como aspas, apóstrofes e etc.) foram removidos, todas as palavras compostas combinadas por um *underline* (por exemplo, Minas_Gerais) e as siglas padronizadas (45). O processo básico de inserção de dados no software pode ser visto na Figura 2.

Figura 2 – Processo básico de inserção de dados no IRaMuTeQ.



Fonte: Esquematização realizada pelo autor a partir do funcionamento do Software IRaMuTeQ.

Antes de detalhar o funcionamento de cada recurso, é importante definir determinados conceitos acerca de sua utilização.

Neste estudo, o software foi empregado para auxiliar na análise da literatura científica bioética sobre o uso de Big Data em saúde. Cada artigo científico sob análise é considerado um texto para o software e o conjunto de todos os textos é chamado de *corpus*. Para realizar suas análises, o IRaMuTeQ divide os textos em segmentos de texto (STs) – que são recortes do texto inteiro com tamanho definido pelo usuário, podendo ser feitos pela contagem de caracteres ou palavras (o padrão é que cada ST possua cerca de 40 palavras). As palavras são chamadas de formas, sendo cada palavra uma forma diferente. Além disso, a lematização é o processo de reduzir uma palavra a seu lema, ou seja, o radical essencial que guarda significado das palavras, por exemplo, as palavras “completar”, “completo” e “completude” são todas reduzidas para o lema “completar”. Para isso os verbos são convertidos ao infinitivo, os substantivos, no singular e os adjetivos, ao masculino singular (45).

O software permite que diferentes tipos de diagnóstico sejam obtidos sobre o *corpus*. Para os fins deste estudo, foram utilizados os recursos de Análise de Classificação Hierárquica Descendente e de Análise de Especificidades (44).

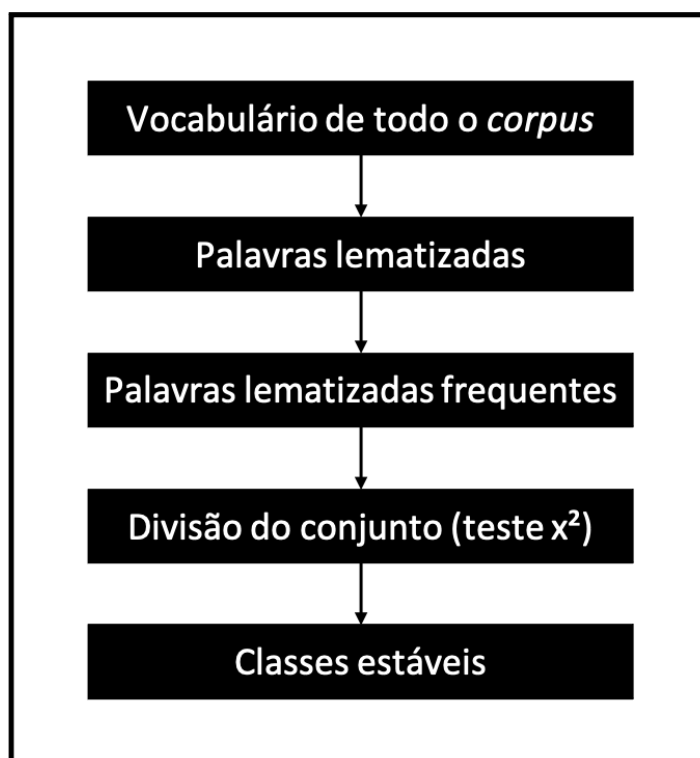
4.3.2 Análise de Classificação Hierárquica Descendente

O método de Classificação Hierárquica Descendente, também conhecido como Método de Reinert, é um método que categoriza os STs de acordo com as semelhanças e diferenças entre si. Após isso, fornece como resultado dessa classificação, classes com STs similares entre si (46).

Com essa finalidade, o método lista todo o vocabulário do *corpus* e depois gera duas listas, uma com as palavras lematizadas, e outra com as mais frequentes dentre elas. Distinguindo palavras com função apenas sintática, ou seja, artigos e conectivos, daquelas que carregam conteúdo, a ferramenta de processamento de dados seleciona as formas, já lematizadas, que aparecem ao menos três vezes (47).

O passo seguinte consiste na categorização de palavras. A primeira divisão de palavras é constituída pelo conjunto total de STs numa tabela com 0 ou 1 que indica se determinada palavra está presente ali. O objetivo do próximo passo é conseguir uma divisão dessa primeira divisão em duas outras, de tal modo que as duas divisões contenham diferenças vocabulares significativas. Para isso, o software testa todas as possibilidades de categorização e vê quais divisões têm o maior valor do χ^2 (quadrado de χ) – que é a fórmula matemática usada para buscar o ponto máximo de similaridade – e aí do presente grupo oferece duas divisões com maior valor do χ^2 . Isso é feito de forma iterativa, até não encontrar mais nenhuma divisão significativa possível, ou seja, quando o software atinge classes estáveis. Com as classes divididas, o programa computa, para cada classe, uma lista de palavras que são características dessa classe. A força da associação entre cada palavra e sua classe é expressa por um valor do χ^2 , e todas as palavras que atingem determinado valor do χ^2 , são listadas. Quanto maior o valor, mais importante é a palavra para a construção estatística da classe (48). Tais etapas podem ser vistas na Figura 3.

Figura 3 – Etapas da divisão de classes de palavras.



Fonte: Esquematização realizada pelo autor a partir do funcionamento do algoritmo utilizado pelo IRaMuTeQ para a divisão de classes de palavras.

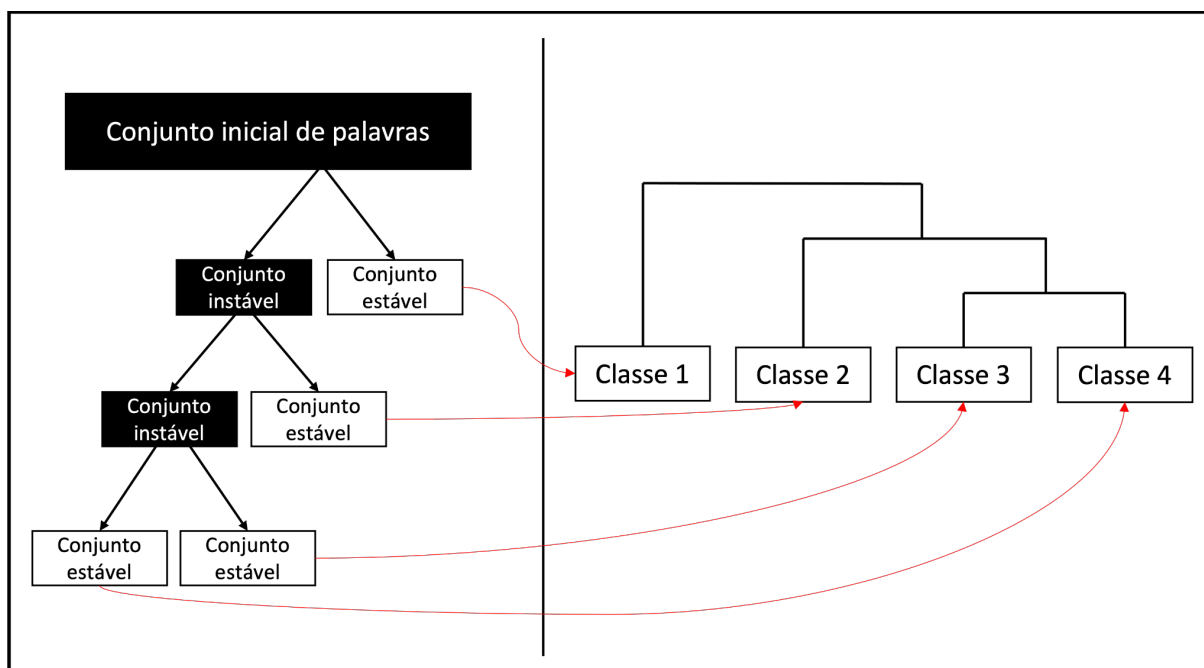
Como resultado da análise realizada no software, são obtidos os chamados dendrogramas, representações gráficas desse processo de categorização textual por similaridade (45). No capítulo de resultados, é apresentado um dendrograma contendo a representação gráfica da divisão de palavras desse trabalho (ver Figura 8).

Além dos dendrogramas, também é apresentada outra representação gráfica, gerada pela Análise Fatorial de Correspondência (AFC), que é uma análise cruzamento entre o vocabulário do *corpus* e as classes. Essa representação gráfica é exibida num plano cartesiano com as palavras de cada classe opostas entre si, dispostas conforme suas semelhanças e diferenças (45). Em outras palavras, é gerado um gráfico com todas as classes agrupadas, o que possibilita avaliar convergências e divergências entre elas. No capítulo de resultados, é apresentado um gráfico resultante da AFC deste trabalho (ver Figura 9).

De forma resumida, o programa lista as palavras do *corpus* e divide esse conjunto inicial de palavras em dois novos conjuntos. Esse processo é repetido até que as palavras em cada conjunto contenham conteúdos significativamente diferentes

entre si, formando classes estáveis. O processo se repete até que as diferenças permaneçam estatisticamente significativas conforme pode ser visto na Figura 4.

Figura 4 – Representação do processo de classificação hierárquica descendente operada pelo IRaMuTeQ.



Fonte: Esquematização realizada pelo autor a partir do funcionamento do algoritmo utilizado pelo IRaMuTeQ para a classificação.

Com esse recurso, o software é capaz de agrupar palavras com conteúdo similar entre si, criando categorias, o que auxilia o entendimento dos textos. Tal tipo de análise evidencia temáticas e tendências em comum, por exemplo, separando o *corpus* entre uma classe que trata de preocupações de natureza prescritiva ou regulatória de outra que remete a temas sociais.

4.3.3 Análise de Especificidades

O recurso de análise de especificidades permite identificar associações entre o texto presente no *corpus* com variáveis definidas pelo pesquisador, de modo que se torna possível detectar tendências relacionadas a elas (46).

Uma vez definidos esses parâmetros, é possível comparar variações em relação a um determinado tema do texto, permitindo, por exemplo, identificar se autores de uma determinada área de conhecimento ou país tendem a remeter mais a um conjunto de palavras do que outras, convergindo em temas e preocupações.

Ao fim da análise de classificação, o software dá como resultado um número determinado de classes de palavras agrupadas por similitude, com apenas os dados quantitativos acerca de cada conjunto de STs.

4.4 IDENTIFICAÇÃO DOS DESAFIOS PARA POSTERIOR ANÁLISE COMPARATIVA

Para realizar a análise comparativa foram analisados os desafios listados no Relatório do IBC sobre Big Data e Saúde (4) e os desafios identificados pelo com auxílio do software na literatura científica selecionada.

Em um primeiro momento, o IRaMuTeQ fornece as classes de palavras similares entre si. Depois, a classificação de cada ST do *corpus* nas classes. Estes STs, uma vez classificados, foram relidos para a identificação dos desafios éticos relacionados ao uso de Big Data em saúde, conforme debatidos na literatura científica. Por vezes, foi necessário remeter à leitura dos próprios artigos, para compreender os aspectos éticos em contexto.

Um exemplo dessa classificação é o segmento de texto: "*current privacy rules do not deal with longitudinal data and transactional data which can be used to reidentify an individual[.] personal genome data are also not covered despite that HIPAA clearly states the protection of biometrics like finger and voice prints*". Esse trecho é identificado pelo IRaMuTeQ como pertencente a Classe 4 (denominada "dimensões regulatórias", como será detalhado no capítulo de Resultados). Conforme análise, esse ST, aqui apresentado como exemplo, remete ao desafio "Privacidade e confidencialidade" (também a ser detalhado no capítulo de Resultados), o que leva este desafio a estar presente na Classe 4.

Conjugando os resultados apresentados pelo software a respeito do *corpus* analisado às demais informações recolhidas dos artigos selecionados, foram identificadas as principais tendências do presente debate sobre o tema, bem como as

convergências e divergências a respeito das possíveis implicações éticas do uso de Big Data no campo da saúde humana, conforme identificadas no Relatório do IBC sobre Big Data e Saúde (4).

5 RESULTADOS

Neste capítulo é apresentado o resultado da análise de moralidade do uso de Big Data em saúde. Conforme descrito na metodologia, o estudo é baseado em uma análise comparativa entre os desafios identificados no [1] Relatório do IBC sobre Big Data e Saúde (4) e os desafios identificados na literatura científica sobre o tema. A seguir, são descritos os resultados referentes às características do processo de seleção do *corpus* analisado [2] e os resultados referentes às análises auxiliadas por recursos do software IRaMuTeQ [3]. Por fim, são apresentados os aspectos éticos relacionados ao uso de Big Data em saúde identificados na literatura científica temática [4].

5.1 RELATÓRIO DO IBC SOBRE BIG DATA E SAÚDE

Definido como base comparativa desta pesquisa, o Relatório do IBC sobre Big Data e Saúde (4) é um documento construído ao longo de dois anos por bioeticistas de todo o mundo para debater o tema de Big Data em saúde, no qual foram abordadas, entre outras, questões relacionadas à autonomia, ao consentimento, à proteção e governança das bases de dados (4).

O IBC descreve na quarta seção do Relatório os principais aspectos teóricos e éticos relacionados ao uso de Big Data em saúde (4).

A Tabela 1 apresenta os aspectos éticos identificados no Relatório do IBC sobre Big Data e Saúde (4).

Tabela 1 – Lista de desafios do Relatório do IBC sobre Big Data e Saúde (4).

Desafio
A. Autonomia e consentimento
B. Privacidade e confidencialidade
C. Da propriedade à custódia e compartilhamento de benefícios
D. Justiça e a lacuna digital
E. Justiça e a não discriminação
F. Sustentabilidade em relação à energia e ao meio ambiente
G. Pesquisa

Fonte: Esquematização realizada pelo autor a partir dos tópicos de desafios que constam no Relatório do IBC sobre Big Data e Saúde (4).

No documento, os aspectos éticos do uso de Big Data em Saúde são divididos nas seções de Autonomia e consentimento; Privacidade e confidencialidade; Da propriedade à custódia e compartilhamento de benefícios; Justiça e a lacuna digital; Justiça e a não discriminação; Sustentabilidade em relação à energia e ao meio ambiente; e, Pesquisa. Apesar de amplo, o próprio IBC ao apresentar o relatório define que este não pretende ser exaustivo (4).

A. Autonomia e Consentimento

O primeiro desafio concerne a **autonomia e consentimento** na aplicação das novas tecnologias informacionais de Big Data em saúde. De acordo com o relatório, o respeito à ação autônoma pressupõe como condições necessárias: a competência para entender e avaliar as informações relevantes; o acesso à informação; possibilidade de escolha entre diferentes opções; a consideração da variação nos valores individuais nas decisões; a voluntariedade das decisões; a garantia de que indivíduos sejam capazes de escolher os objetivos e meios para suas ações; e, a ação em si tem de ser algo consciente. A autonomia de um indivíduo relaciona-se a sua capacidade de tomar uma decisão e, tradicionalmente, principalmente no âmbito da pesquisa, o consentimento se apresenta como sua principal forma de expressão e garantia. O tema da autonomia, recorrente nas discussões em bioética e em ética em pesquisa, apresenta especificidades que precisam ser observadas durante a reflexão ética a respeito do uso de Big Data em saúde (4).

O respeito à autonomia e a obtenção de consentimento dos indivíduos são dificultados pela escala e a complexidade das análises e processamentos de dados realizados em Big Data. O acompanhamento do uso e apropriação das informações

e a tomada de decisões diante disso se tornam virtualmente impossíveis à medida que as bases cruzam cada vez mais dados. O resultante geral do processo é um decréscimo da autonomia quanto ao uso dos dados para finalidades e autores diversos do que se consentiu a princípio (4).

Os meios de obtenção do consentimento apresentam uma série de dificuldades. Devido à forma como são concebidos, ou seja, como termos de adesão, existe pouca margem para real discordância. Adicionalmente, como são obtidos por vias eletrônicas, não há maneiras de verificar se os indivíduos estão consentindo em suas plenas capacidades, ou se são maiores e capazes legalmente de fazê-lo (4).

Além disso, a imprevisibilidade, característica da tecnologia, se apresenta como um fator complicador diante daquilo que compreendemos como condições para tomadas de decisão. A possibilidade de intersecção dos dados e acréscimos imprevistos nas fontes das bases de dados e usos futuros imprevistos colocam em pauta qual seria a forma adequada de consentimento nesse contexto, de modo que ainda que o consentimento seja uma etapa no uso das tecnologias, existem dúvidas quanto a se isso de fato é um consentimento livre e informado (4).

Diversas abordagens são propostas para obtenção do consentimento, no entanto, até o momento, nenhuma dessas soluções parece ser completa ou definitiva em virtude da incerteza característica dessas tecnologias. Um exemplo é o consentimento amplo ("*broad consent*"). O consentimento amplo é um instrumento concebido para ser flexível quanto ao seu escopo, porque o uso dos dados passa a ser descrito de forma mais vaga e abrangente, operando em um nível mais alto de abstração (4). Por exemplo, no lugar de um paciente ceder seus dados e consentimento para diferentes pesquisas de câncer diversas vezes seguidas, por meio do consentimento amplo ele pode ceder os dados e o consentimento apenas uma vez, delimitando que qualquer pesquisa para o tratamento de câncer possa utilizar seus dados. Outra proposta de abordagem é o "*opt-out*", ou seja, uma forma negativa de consentimento em que se trabalha com a hipótese de que os dados de saúde podem sempre ser usados para pesquisa a menos que um indivíduo expresse seu desejo de não participação (4). Todas as opções, no entanto, não reestabelecem plenamente a capacidade do indivíduo de decidir sobre o uso de seus dados pessoais. De tal modo, o IBC reconhece que o consentimento é um grande desafio, especialmente no uso de aplicativos de saúde e na área de mídias e redes sociais. Ao apontar tais dificuldades, o Relatório destaca o papel da educação pública, de

modo que esta contribua para que usuários, pacientes e participantes de pesquisa entendam o alcance e impacto dos possíveis usos de seus dados (4).

B. Privacidade e confidencialidade

A questão da **privacidade e confidencialidade** figura como outro desafio central relacionado ao uso de Big Data. O direito à privacidade é fortemente conectado ao direito à liberdade com seus diversos aspectos éticos e legais como liberdade de expressão, de associação, de localização, de movimento e espaço, de crenças, pensamentos e sentimentos, e de comportamento. Com esse contexto, o IBC usa o termo privacidade no sentido de um direito ao respeito a uma vida particular ou aos dados que os indivíduos querem manter reservados para si mesmos ou, ao menos, para membros específicos de suas famílias ou relacionamentos. Seguindo as diretrizes da DUBDH, a privacidade e confidencialidade de um indivíduo são um direito (4).

Muitas das técnicas utilizadas ao longo do tempo para a proteção dos dados não se aplicam ao uso de Big Data, já que este frequentemente implica mudança e abertura de propósito, falta de limites para os dados, processamento não transparente, pouca proteção dos dados para ganhar o máximo de conhecimento possível e não transparência para indivíduos envolvidos como características fundamentais. De tal modo, o uso de Big Data desafia as técnicas e mecanismos antes efetivos na proteção dos indivíduos (4).

Ainda, um dos principais interesses na utilização de Big Data reside justamente na utilização massiva de dados agregados, de maneira diversa dos propósitos iniciais da coleta. Os dados não são limitados e são cruzados para inferir novas correlações que não estavam previstas inicialmente, tendo como um dos focos da tecnologia a geração de experiências personalizadas. Todos esses fatores tornam evidente a necessidade de proteção dos indivíduos, ainda que o documento se posicione categoricamente em favor do benefício coletivo da circulação de informações (4).

C. Da propriedade à custódia e compartilhamento de benefícios

O relatório identifica a temática da **propriedade à custódia e compartilhamento de benefícios** relacionados ao seu uso. Para o IBC, o conceito

de propriedade dos dados, em termos de controle, envolveria “o direito dos sujeitos dos dados de impedir qualquer tipo de manipulação de dados...”, o que já não seria possível nos dias atuais segundo o documento (4, p. 19). Já o conceito de custódia envolve “a responsabilidade pela segurança e bem-estar de alguém ou algo e representa valores éticos como cuidado, custódia, proteção e confiança à custódia ou à guarda” (4, p. 15).

O IBC afirma que é cada vez mais consensual na comunidade científica a importância de compartilhar dados científicos de forma mais ampla. Acredita-se que com o compartilhamento de dados, de forma justa e consentida, o conhecimento gerado poderá trazer resultados benéficos para toda a humanidade. Ao mesmo tempo, reconhece que o compartilhamento de dados evoca questões relativas ao consentimento, à posse dos dados, aos direitos de propriedade intelectual e à privacidade. Deste modo, torna-se cada vez mais problemático resolver os desafios éticos e legais utilizando o conceito de propriedade (4).

D. Justiça e a lacuna digital

O relatório também aborda a temática de **justiça e a lacuna digital**. Apesar de ser um grande trunfo no combate à pobreza, o acesso à tecnologia ainda apresenta muita disparidade, tanto entre estratos sociais diversos, quanto entre países periféricos e centrais. Há uma demanda elevada por infraestrutura e recursos humanos para fazer funcionar as ferramentas e provedores necessários a essas tecnologias. Adicionalmente, existe uma grande disparidade entre habilidades tecnológicas, alfabetização digital e a existência de conteúdo local relevante entre países centrais e periféricos (4).

Dada essa diferença de contexto, é comum que exista uma transferência massiva de dados de países periféricos para países centrais e já desenvolvidos tecnologicamente, o que ocorre sem garantias sobre os dados que são transferidos. Assim, a lacuna digital se constitui como um dos maiores desafios para a democratização da informação, comunicação e seu desenvolvimento (4).

É significativo que, neste ponto, o relatório não aponte como essas tecnologias podem atuar como fatores adicionais de concentração de poder e renda, aumentando as mesmas desigualdades que o documento identifica.

E. Justiça e a não discriminação

Ainda ao abordar o tema Justiça, o relatório fala de **justiça e a não discriminação** como um de seus desafios específicos. Essa seção aborda desafios ligados ao uso dos dados e riscos de discriminação com base nestes. O IBC cita, por exemplo, o uso indevido de dados por parte das seguradoras de saúde. Em muitos dos países que não reconhecem a saúde como um direito universal, o acesso a ela, mesmo que pago, pode ser negado por seguradoras de saúde. Condições pré-existentes, tendências genéticas e práticas consideradas de risco podem ser usadas pelas seguradoras como justificativa para a exclusão dos indivíduos. Tais práticas, destinadas a aumentar a lucratividade das operadoras, podem, na vigência do fenômeno Big Data, ser intensificadas por práticas cada vez mais invasivas de investigação, com uso de informações pessoais vazadas ou até mesmo lançando mão do processo de classificação de indivíduos em perfis de risco. A discriminação e o agravamento de injustiça contra grupos já vulnerabilizados é uma preocupação importante (4).

F. Sustentabilidade em relação à energia e ao meio ambiente

Outro desafio apresentado pelo relatório é o da **sustentabilidade em relação à energia e ao meio ambiente**. O elevado poder computacional requer alto consumo de energia elétrica para seu processamento e refrigeração, além da construção de infraestrutura para o seu funcionamento. Adicionalmente, o processo de prospecção de matérias-primas para a produção de componentes essenciais aos equipamentos eletrônicos tem impactos ecológicos significativos (4).

Entretanto, o Relatório aventa a possibilidade de que a tecnologia traga resultados ecológicos positivos, uma vez que as pesquisas que lançam mão de aglomerados de dados podem ser um recurso para a busca por conhecimento a respeito do uso mais eficiente de recursos naturais (4).

G. Pesquisa

Por fim, o desafio da ética em **pesquisa** envolvendo Big Data é também especialmente tratado no Relatório. Dado que os métodos de Big Data podem

constituir poderosos geradores de hipóteses para pesquisa, também esta fase deve observar os mais altos padrões éticos, com a preocupação na proteção de dados e excelência em sua utilização. Já existem, nesse sentido, algumas orientações de melhores práticas na pesquisa com Big Data, que incluem, por exemplo, a criação de comitês com representação de pacientes para a governança de biobancos e sua participação nos processos de acompanhamentos de informações sobre as pesquisas com tais dados (4).

Embora aponte todos esses aspectos éticos, o relatório alerta que posições regulatórias que diminuam a circulação dos dados podem se apresentar como obstáculo para pesquisa e produção do conhecimento. Apesar de mencionar a importância da propriedade dos dados pessoais e dos direitos provenientes destes, o IBC afirma que esses conceitos tradicionais podem ter se tornado inadequados. É necessário, de acordo com o documento, que se desenvolva uma nova estrutura normativa com novos conceitos, de modo a equilibrar os interesses envolvidos e não atrapalhar a produção do conhecimento (4).

Assim, reconhecendo os desafios conceituais e práticos que envolvem as medidas para a necessária proteção dos dados, o IBC aponta para a necessidade de maior reflexão quanto a medidas que ao mesmo tempo garantam a segurança, autonomia e privacidade e viabilizem o compartilhamento de dados para o bem comum (4).

Outro aspecto destacado pelo Relatório que envolve a temática da pesquisa é a necessidade de educação da população, imprescindível para o exercício da autonomia na participação de pesquisas, mas também para que os resultados sejam interpretados de forma correta (4).

Compreendidos os principais aspectos enumerados pelo Relatório, passa-se a tratar a literatura científica sobre o tema.

5.2 CARACTERÍSTICAS DO PROCESSO DE DETERMINAÇÃO DO *CORPUS* DE ANÁLISE DA LITERATURA CIENTÍFICA

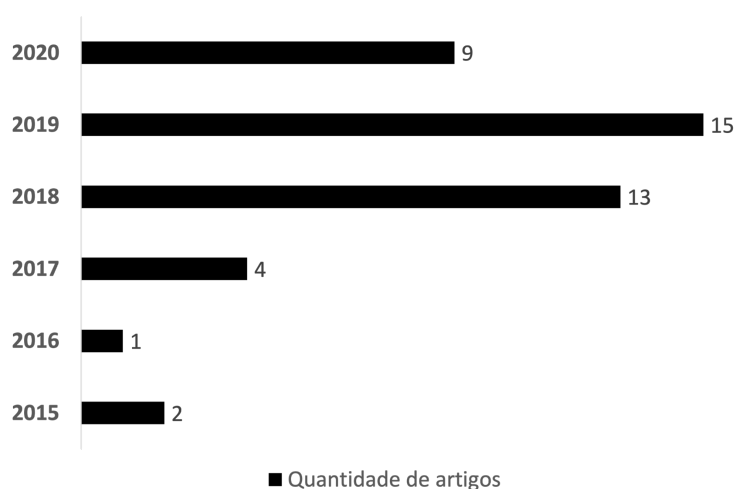
A busca inicial foi realizada em nível nacional, regional e internacional. Em nível regional, a base escolhida foi a LILACS (que por sua vez engloba a base nacional SciELO). Observados os critérios de busca, não houve resultados para a busca sobre a temática Big Data em saúde nos níveis nacional e regional, o que indica incipiência da discussão ética sobre o tema.

Em nível internacional, com a busca na base PubMed, foram encontrados 55 artigos. Após a verificação dos critérios de inclusão e exclusão, um artigo em francês foi removido por estar fora dos idiomas selecionados para a pesquisa e outros 10, por não se encontrarem dentro do escopo temático próprio da pesquisa. De tal modo, o *corpus* da pesquisa foi composto por 44 artigos.

A lista completa dos artigos analisados e demais informações sobre estes encontra-se no Anexo B.

Em relação aos anos em que os artigos foram publicados, conforme visto na Figura 5, percebe-se uma tendência crescente ao longo dos anos, o que indica um aumento de interesse no tema.

Figura 5 – Quantidade de artigos selecionados por ano.



Fonte: Gráfico elaborado pelo autor com base nas informações dos artigos analisados durante a pesquisa.

Em relação aos países de publicação dos artigos, conforme pode ser visto na Figura 6, a produção sobre o tema é principalmente anglo-saxônica, e majoritariamente estadunidense. Dos 44 artigos encontrados, 19 são dos EUA, que totalizam mais do que o dobro do segundo país, o Reino Unido, com 8 artigos.

Figura 6 – Quantidade de artigos selecionados por país.



Fonte: Gráfico elaborado pelo autor com base nas informações dos artigos analisados durante a pesquisa.

Os artigos foram ainda classificados quanto à área de atuação científica de seus primeiros autores das publicações encontradas, conforme Figura 7 abaixo. Foram identificadas cinco áreas, percebe-se, no entanto, a prevalência de autores atuantes na saúde e tecnologia, com 26 das 44 publicações. Ainda, áreas comumente não associadas à bioética, como administração, se fizeram representar na amostra. Tal fato evidencia o grande valor do uso dados para a administração de instituições privadas, que buscam formas mais efetivas e lucrativas de atuação com o auxílio de Big Data. Ainda, em alguns casos, não foi possível identificar a área do primeiro autor com base nas informações presentes no artigo.

Figura 7 – Quantidade de artigos selecionados por área do primeiro autor.



Fonte: Gráfico elaborado pelo autor com base nas informações dos artigos encontrados na pesquisa.

5.3 ANÁLISES REALIZADAS COM O APOIO DO SOFTWARE IRAMUTEQ

Após a verificação dos critérios seguiu-se com a análise de todo o *corpus* da pesquisa apoiada pelo software IRaMuTeQ, com a Análise de Classificação Hierárquica Descendente e Análise de Especificidades.

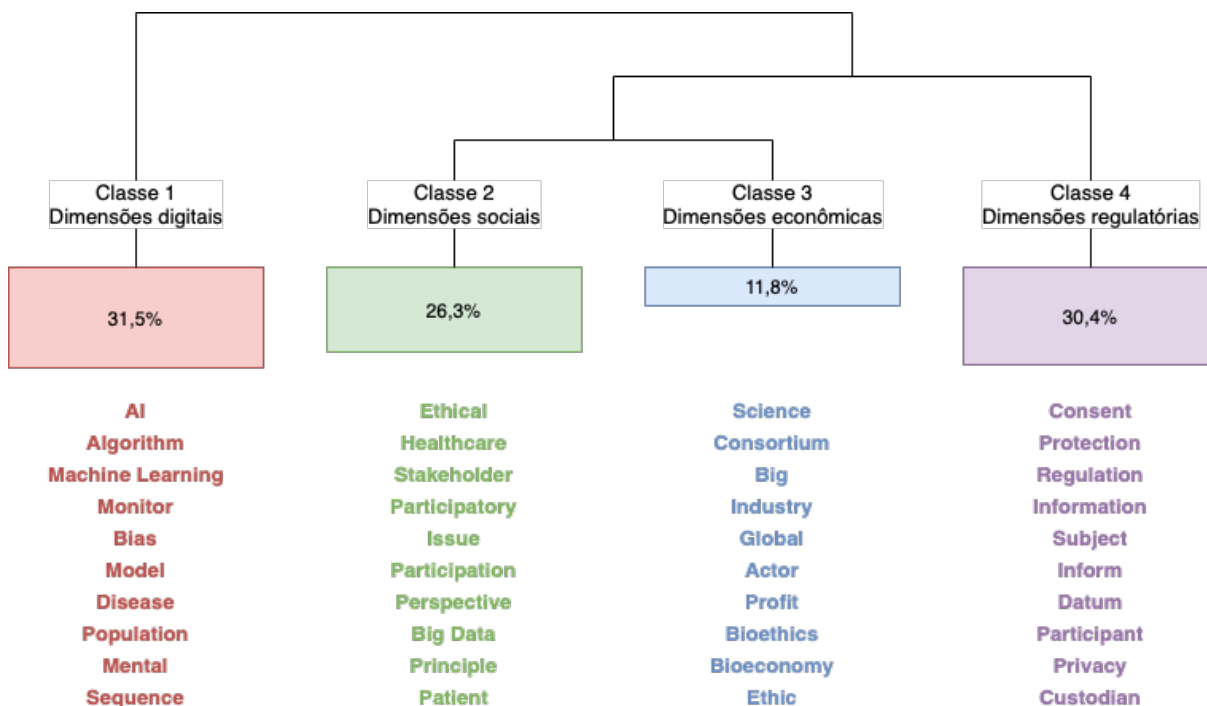
5.3.1 Análise de Classificação Hierárquica Descendente

Na análise de classificação, o software identificou quatro classes de palavras. O conteúdo analisado foi categorizado em quatro classes: Classe 1, com 31,53% do *corpus*; Classe 2, com 26,28%; Classe 3, com 11,76%; e Classe 4, com 30,42%.

Conforme tratado na metodologia, o software lança mão de mecanismos estatísticos para a categorização de palavras semelhantes nas mesmas classes, mas não as nomeia. É necessária uma leitura e análise dos resultados do processamento do software. Desta forma, observando os termos mais frequentes de cada classe em contexto, os nomes das classes refletem as diferentes dimensões da moralidade do

uso de Big Data. Os nomes das classes, conforme definidos pelo pesquisador, constam na Figura 8.

Figura 8 – Dendrograma com as classes geradas e suas principais palavras.



Fonte: Dendrograma criado pelo autor com base no diagrama resultante da análise do IRaMuTeQ.

A seguir, em um primeiro momento, são apresentados detalhes sobre as classes geradas pelo software. Na próxima seção, após a apresentação das classes, os desafios presentes em cada uma delas são apresentados em detalhes. Mais informações quanto aos dados brutos da análise de classificação, fornecidos pelo software IRaMuTeQ, podem ser vistos no relatório de resultado contido no Anexo C.

5.3.1.1 Classe 1 – Dimensões Digitais

Assim nomeada pela presença de diversos desafios de natureza técnica e pela presença de diversos termos técnicos na lista dos termos mais frequentes da classe. Na análise verificou-se que nessa classe estavam presentes STs relacionados aos

desafios no campo propriamente tecnológico da utilização de Big Data. Isso se torna evidente ao analisar as principais palavras da classe.

Compreende 31,5% do *corpus* total analisado. É constituída por palavras e radicais no intervalo entre $\chi^2 = 177,21$ (“*artificial_intelligence*”) e $\chi^2 = 3,92$ (“*traditional*”). É uma classe composta por palavras como “*algorithm*” ($\chi^2 = 167,55$); “*machine_learning*” ($\chi^2 = 135,56$); “*monitor*” ($\chi^2 = 114,15$); e “*bias*” ($\chi^2 = 114,05$).

5.3.1.2 Classe 2 – Dimensões Sociais

Remete a uma percepção coletiva das dimensões morais do uso de Big Data em saúde. Na análise verificou-se que o conteúdo da classe envolve STs relacionados a saúde e participação da sociedade em tais temas, de modo que desafios que impactam a sociedade e grupos dentro dela se mostraram evidentes dentro da mesma.

Compreende 26,3% do *corpus* total analisado. É constituída por palavras e radicais no intervalo entre $\chi^2 = 389,06$ (“*ethical*”) e $\chi^2 = 3,92$ (“*economic*”). É uma classe composta por palavras como “*healthcare*” ($\chi^2 = 118,06$); “*stakeholder*” ($\chi^2 = 103,23$); “*participatory*” ($\chi^2 = 98,59$); e “*issue*” ($\chi^2 = 91,13$).

5.3.1.3 Classe 3 – Dimensões Econômicas

Engloba os desafios relacionados às relações entre indivíduos fornecedores de dados e entes privados que se utilizam destes, possui a presença de termos como “lucro”, “bioeconomia” e “indústria”. Na análise verificou-se a presença de STs relacionados aos desafios geralmente ligados à comercialização de dados e à participação de corporações na coleta destes.

Compreende 11,8% do *corpus* total analisado. É constituída por palavras e radicais no intervalo entre $\chi^2 = 588,14$ (“*science*”) e $\chi^2 = 3,85$ (“*dollar*”). É uma classe composta por palavras como “*consortium*” ($\chi^2 = 489,62$); “*big*” ($\chi^2 = 167,31$); “*industry*” ($\chi^2 = 137,06$); e “*global*” ($\chi^2 = 124,82$).

5.3.1.4 Classe 4 – Dimensões Regulatórias

Classe que envolve os desafios ligados à regulação da tecnologia, propriedade dos dados e proteção dos indivíduos. Na análise verificou-se STs relacionados à discussão de qual é o melhor tipo de consentimento e se a regulação atual é ou não o suficiente para lidar com a crescente demanda por soluções de proteção jurídica.

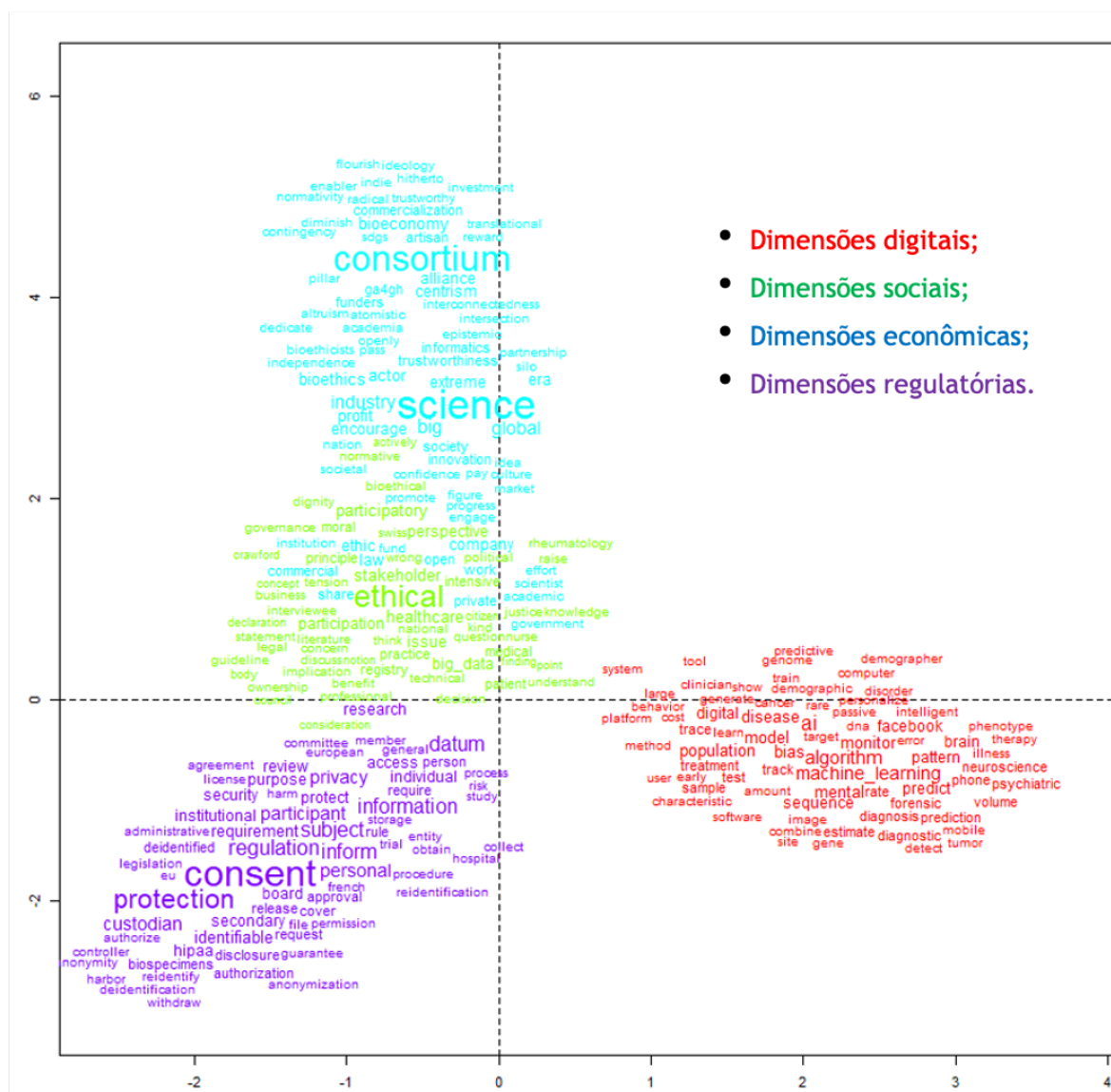
Compreende 30,4% do *corpus* total analisado. É constituída por palavras e radicais no intervalo entre $\chi^2 = 505,53$ (“*consent*”) e $\chi^2 = 3,9$ (“*readily*”). É uma classe composta por palavras como “*protection*” ($\chi^2 = 309,42$); “*regulation*” ($\chi^2 = 206,93$); “*information*” ($\chi^2 = 192,39$); e “*subject*” ($\chi^2 = 189,98$).

5.3.1.5 Análise Fatorial por Correspondência

Ainda na análise de classificação, a partir da AFC, foi possível realizar associação do texto entre as palavras, considerando a frequência de incidência entre as palavras e as classes, representando-a num plano cartesiano, como é possível ver na Figura 9.

Observa-se que a maior aproximação entre os discursos se dá entre as classes 2 (Dimensões sociais) e 3 (Dimensões econômicas). As palavras da Classe 1 (Dimensões digitais) se apresentam agrupadas entre si, mas isoladas das demais classes. A Classe 4 (Dimensões regulatórias) se encontra mais próxima das classes 2 (Dimensões sociais) e 3 (Dimensões econômicas) do que aquelas da classe 1.

Figura 9 – Representação num plano cartesiano das palavras de cada classe.



Fonte: Exportado da análise do IRaMuTeQ pelo autor.

5.3.2 Análise de Especificidades

Como detalhado no capítulo que trata da metodologia, a análise de especificidades permite comparar o texto com variáveis definidas previamente. Dois tipos de variáveis foram utilizados, uma sendo o país de publicação dos artigos e outra a área dos primeiros autores.

Quanto aos países de publicações dos artigos, na Classe 1 (Dimensões digitais), predominaram artigos dos Estados Unidos, Turquia e Brasil; na Classe 2 (Dimensões sociais), predominaram artigos da Alemanha, Suíça e Austrália; na Classe 3 (Dimensões econômicas), predominaram artigos do Reino Unido e da Áustria; e, na Classe 4 (Dimensões regulatórias), predominaram artigos da Espanha, Austrália e França. Uma análise preliminar reflete uma preocupação diversa conforme o país. Essa diferença pode ser atribuída a especificidades do contexto sociopolítico nacional.

Quanto às áreas de atuação acadêmica dos primeiros autores de cada artigo, na Classe 1, dimensões digitais, predominaram artigos das áreas de saúde e tecnologia; na Classe 2, dimensões sociais, predominaram artigos da área de ética, bioética e administração; na Classe 3, dimensões econômicas, predominaram artigos da área de direito e administração; e, na Classe 4, dimensões regulatórias, predominaram artigos das áreas de direito e tecnologia. Analisando de forma preliminar, a predominância de artigos de diferentes áreas nas classes tende a expressar preocupações gerais daquela área no contexto do uso de Big Data, como, por exemplo, a presença da área de tecnologia na classe de dimensões digitais se explica devido ao interesse técnico do tema para a área.

Assim como os dados acerca da análise de classificação, os dados brutos da análise de especificidades, fornecidos pelo software IRaMuTeQ, podem ser vistos no Relatório constante no Anexo C.

5.4 IDENTIFICAÇÃO DOS ASPECTOS ÉTICOS NA LITERATURA CIENTÍFICA RELACIONADOS AO USO DE BIG DATA EM SAÚDE

Após a identificação das classes de aspectos éticos e leitura dos artigos presentes no *corpus*, todos os aspectos relacionados à utilização de Big Data em saúde identificados foram registrados e os segmentos de textos foram então categorizados, possibilitando identificar o local em que cada aspecto era mencionado e quais deles cada artigo levantava.

Ainda, os aspectos foram enquadrados dentro de cada uma dessas classes, de acordo com seu tema.

Na Tabela 2, eles são listados em suas devidas classes.

Tabela 2 – Lista de aspectos separados por classes.

Classe	Aspectos éticos
Classe 1 – Dimensões digitais	Condição humana
	Incerteza e imprevisibilidade
	Cooperação
	Opacidade
	Segurança digital
	Qualidade e representatividade
Classe 2 – Dimensões sociais	Apropriação do conhecimento
	Criação de perfis e autonomia
	Conflito entre interesses individuais e coletivos
	Injustiça e desigualdades
	Modulação comportamental
	Preconceito, estigma e injustiça
	Tomada de decisão baseada em dados e responsabilidade moral
Validade do conhecimento	
Classe 3 – Dimensões econômicas	Comoditização da privacidade
	Percepção pública
Classe 4 – Dimensões regulatórias	Consentimento
	Mecanismos de proteção
	Privacidade e confidencialidade

Fonte: Esquematização realizada pelo autor a partir das classes geradas pelo software IRaMuTeQ e os aspectos éticos sobre o uso de Big Data em saúde identificados na literatura.

A seguir, cada aspecto/desafio encontrado é descrito e apresentado junto a trechos que o exemplificam.

5.4.1 Classe 1 – Dimensões Digitais

Nas dimensões digitais, foram identificados como aspectos éticos no *corpus* analisado: condição humana, incerteza e imprevisibilidade, cooperação, opacidade, segurança digital, e qualidade e representatividade.

A. Condição humana

Um dos desafios relatados por alguns autores é que a discussão presente na literatura científica ainda não tem se atentado suficientemente como essa transformação digital atua sobre a compreensão do que é a condição humana:

Right now, the discussion revolves around data security, but it barely tackles the shift that constitutes the human being as a digital being and hence affects scientists' understanding of the human body as well as of the human condition and the research questions that are pursued (49, p. 377).

Em contraste, há questionamentos quanto à capacidade descritiva dos fenômenos a partir de Big Data. Dessa maneira, argumenta-se que os processos decisórios pautados por Big Data podem não ter acesso à variedade e profundidade das dimensões da realidade humana. Qualquer resultado, por mais completo que seja, é apenas uma parte daquilo que tenta descrever e as extrapolações podem ter consequências desastrosas. Um exemplo dessa preocupação é expresso nos seguintes trechos destacados da análise:

Big Data provides a worldview that is a reflection of everyone but of no one. Decisions about many aspects of life, from buying habits to technology usage, may soon be predicated on these massive data sets. But we question whether the depth, breadth, and variety of information necessary to make critical decisions are captured by this artificial reality? The revealed truth may not be the whole truth but rather a shadow of truth (50, p. 79).

... using automated text analysis techniques on large textual data provides only a partial understanding of the meaning embedded within this content. And although analyzing word counts or using machine learning techniques (i.e., methods in which algorithms are trained to recognize user traits based on patterns inferred from input data) to categorize text may provide a rough illustration of ideas covered within a corpus of data, they do not provide a deep understanding of the processes that generated them or of user intent (51, p. 6).

B. Incerteza e imprevisibilidade

A incerteza e a imprevisibilidade emergem como desafio/questão ética na classe dimensões digitais. O uso de Big Data visa justamente identificar padrões de comportamento não previstos, obtidos a partir da conjugação e entrecruzamento de dados de diferentes fontes, tamanhos e aspectos, que por sua vez nem no momento da coleta são passíveis de total conhecimento sobre seu uso.

Tendo em vista essa característica, é virtualmente impossível determinar toda a gama futura de riscos (e também benefícios) de uma pesquisa e/ou análise de antemão:

Since it is hard to tell which data will be useful in future research... (52, p. 2).

Além disso, também pode não ser possível identificar quais dados serão reutilizados em algum momento futuro, já que seus potenciais usos e readaptações são virtualmente ilimitados. Essa incerteza coloca limites à autonomia de quem fornece os dados, tendo em vista a impossibilidade de tomar decisões com base em tudo o que potencialmente pode ser feito com tais dados:

Likewise, given recent evolution in how individuals view and understand digital privacy, it may be difficult to assess the risks and benefits of research that uses these novel data sources until after the research is conducted or published (51, p. 7).

C. Cooperação

Ainda na classe de dimensões digitais, emergiu a questão/dimensão ética da cooperação no acesso a dados. Considerando o valor estratégico informacional, observam-se tendências de monopolização de acesso aos dados tanto por parte de governos quanto entidades privadas. Em determinados casos, inclusive por razões de segurança, é bom que bancos de dados de informações pessoais estejam isolados e não compartilhando informações. Todavia, em casos nos quais a sociedade como um todo pode se beneficiar do compartilhamento de informações (devidamente consentidas e com as medidas de segurança necessárias aplicadas), a monopolização do acesso é algo maléfico, indo contra o compartilhamento de benefícios oriundos dos dados. Os trechos a seguir são exemplos de maneiras com que o problema é levantado:

One of the challenges that can undermine effective linking of multiple sources of information is when the data repositories that would benefit from being interlinked and appropriately mined to deliver added value, instead sit in siloed isolation with no ability to 'talk' to each other (53, p. 67).

To prevent data-siloing and ensure that all can benefit from these technologies, we need to start treating geno-/phenotypic and other digital health data as public goods rather than private resources (54, p. 273).

D. Opacidade

Uma das questões/aspectos éticos de maior significado na classe denominada de dimensões digitais é a opacidade. As análises de Big Data ocorrem por meio de algoritmos que muitas vezes envolvem aprendizado de máquina. Apesar de inicialmente programado, o aprendizado de máquina, como o próprio nome sugere,

utiliza-se de recursos que permitem que o algoritmo vá além de seu formato inicial, de modo que apesar de originalmente desenvolvido por humanos, ele evolui de forma autônoma. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

... users do not explicitly know how the output is being obtained given the complexity of models with various parameters and the vast amount of data inputs. In many cases, this is not because of a lack of expertise or knowledge, but rather because AI carries out processes difficult to understand and explain in human terms, and can reach unexpected conclusions (55, p. 1010).

Tendo em vista a natureza de tais algoritmos, é comum ao analisar um resultado de uma análise de Big Data, os resultados saíam de “uma caixa preta”, ou seja, que os resultados encontrados sejam resultantes de processos desconhecidos. Os resultados dessas análises não são auditáveis, o que traz implicações importantes para a tomada de decisões:

... it is often unclear how decisions are made (50, p. 80).

Artificial intelligence and machine learning can be difficult to trust, because in some cases, the algorithms rely on intricate, difficult-to-understand mathematics to get from data input to the final result, resulting in a black box effect (56, p. 3).

E. Segurança digital

Tratando das dimensões digitais, talvez a preocupação mais evidente entre todas aquelas relacionadas ao uso de dados pessoais, dados sensíveis e de uma enorme quantidade de indivíduos seja a segurança no armazenamento de todos esses dados. Não são incomuns as notícias de vazamentos de dados de milhões de indivíduos por corporações dos mais variados segmentos. Somam-se a elas os relatos de governos e indivíduos publicamente expostos. Assim, é evidente o tamanho do desafio e responsabilidade para gestores de bancos de dados, especialmente os que possuem informações capazes de gerar discriminação e preconceito em caso de brechas. Mesmo procedimentos como a encriptação e anonimização podem não ser medidas bastantes, conforme descrevem os seguintes trechos:

... the risk of security breaches may be higher in settings with lower levels of literacy, higher levels of corruption, or which are undergoing rapid technological transition without oversight (55, p. 1012).

An issue, which cannot be ignored, is possible security breaches, as for instance even with encryption and controlled authorized access only, a secure virtual workplace and network might not exclude all types of data transfers, such as, for example, printing, downloading, and e-mailing (49, p. 374).

Policy makers and scholars alike recognize that the public and patients need to know that their data is protected, safe and won't lead to personal harm (57, p. 10).

F. Qualidade e representatividade

Como dimensão digital, o último e um dos principais aspectos identificados durante a análise se refere à qualidade dos dados utilizados para alimentar os algoritmos e os próprios algoritmos que realizam as análises. A origem dos dados pode influenciar os resultados, tendo em vista que dados de “origens obscuras”, coletados de forma desorganizada ou mal estruturada, ou de fontes inverídicas geram resultados não confiáveis. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

The tendency to rely on mere Big Data furthermore ignores the variable quality of datasets (1, p. 320).

Outro grande desafio quanto ao uso dos dados é a falta de representatividade. Os bancos de dados podem possuir processos que garantam a qualidade da coleta até a análise, entretanto, caso não haja preocupação com vieses durante todo o processo de coleta e uso dos dados, a falta de representatividade pode contribuir para discriminação e/ou falta de atendimento/cuidado para determinados grupos não representados. Esse mecanismo pode inclusive agir como retroalimentação de distorções sociais existentes, contribuindo para a aceleração da lacuna digital que representou o ponto de partida. Esse entrecruzamento entre as dimensões digitais e mecanismos de concentração de poder foram apreendidos em trechos como os seguintes:

... Big Data leads to new concentrations of power, 'blind spots' and problems of representativeness because it “cannot account for those who participate in the social world in ways that do not register as digital signals (1, p. 324).

Another ethical concern raised in the integration of registries is dataset representativeness (55, p. 1009).

We expand on this argument to suggest that this discovery/delivery disconnect leads to a less obvious but potentially more powerful ethical issue of systematic absence of particular groups from the data used to develop such technologies, often those groups most disfavored by disparities (58, p. 87).

5.4.2 Classe 2 – Dimensões Sociais

Nas dimensões sociais, foram identificados como aspectos éticos no *corpus* analisado: apropriação do conhecimento; criação de perfis e autonomia; conflito entre interesses individuais e coletivos; injustiça e desigualdades; modulação comportamental; preconceito, estigma e injustiça; tomada de decisão baseada em dados e responsabilidade moral; e validade do conhecimento.

A. Apropriação do conhecimento

Com o uso de Big Data, existe o risco de derivar relações causais injustificadas de fenômenos meramente correlacionados, de modo que a qualidade da análise se torne não satisfatória. Como não é possível auditar todo o caminho que foi percorrido ao gerar determinado resultado, é possível que conclusões equivocadas e enviesadas passem a ser tomadas como verdadeiras. Isso poderia ocorrer mesmo que para o público e até mesmo para *experts* fosse difícil comprovar problemas com a validade do conhecimento produzido. Exemplos dessa preocupação são expressos nos seguintes trechos encontrados na literatura:

Harm can occur to the data subject through misinterpretation or overreliance on data representing the subject's state(s) of being (1, p. 328).

Another threat is the fact that data analysis is somewhat cryptic for most of us as understanding underlying assumptions, algorithms, and the predictive nature of calculating probabilities and correlations rather than causality from large data sets easily slip the mind, which can lead to misconceiving data as facts (49, p. 375).

Ainda, para avaliar e utilizar as tecnologias, sempre em expansão, a formação adequada para a interpretação das informações por vezes é rara ou mesmo inexistente. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

However, there is neither a specific professional degree or certification program for designating who is an expert at rendering health information de-identified particular, nor a method for assessing the level of risk and assuring that it meets the “very small” level indicated by the method (59, p. 109).

B. Criação de perfis e autonomia

Quanto à autonomia, o *corpus* aponta que é possível que indivíduos não consigam mais proteger seus próprios interesses. O procedimento de criar perfis de comportamento utilizados para prever características e tendências de grupos acaba por afetar a autonomia e privacidade mesmo de eventuais indivíduos que consigam permanecer em “um isolamento total”. A criação de perfis é uma questão extremamente desafiadora. Indivíduos sem contato com a análise ou a coleta de dados podem possuir características demográficas comuns ou similaridades genéticas com perfis de grupos identificados. Desse modo, mesmo a coleta de dados anonimizados podem ter como consequência a discriminação num nível individual.

O uso de Big Data constitui um problema de dimensões sociais. O enfoque em perspectivas individualistas de solução dos desafios, como a anonimização de dados e a autonomia e consentimento dos indivíduos, é inócuo. Adicionalmente, acaba por desmobilizar e prejudicar o enfrentamento das questões éticas em suas reais proporções sociais, conforme ilustram os seguintes trechos:

In contrast, the critique grounded in privacy interdependence highlights an altogether different problem: individuals cannot, through their own autonomous decision making, protect their own interests anymore. No person is an island in the environment of Big Data. Insistence on atomistic autonomy disempowers individuals, if the problems they face require collective action (60, p. 5).

... AI-powered neurotechnologies can still affect personal autonomy by having an impact on their cognitive liberty (61, p. 84).

... group-level harms from analysis of aggregated data are clearly possible (1, p. 318).

It is possible to discriminate even with perfectly anonymized data, since predictions about a group affect all members regardless of individual identities (55, p. 1011).

... genomic data may disclose information about people beyond the individual from whom the data were sampled (62, p. 301).

C. Conflito entre interesses individuais e coletivos

Um dos principais desafios ligados a sociedade é o de balancear os benefícios trazidos pelo compartilhamento de dados para a sociedade com os danos e riscos que podem afetar o indivíduo:

On the other hand, there is a constant conflict between the benefit of using multiple sources of information and the value of preserving confidentiality of medical information (59, p. 91).

O avanço do conhecimento pode depender de dados, entretanto, isso não pode acontecer às custas de indivíduos, principalmente daqueles que se encontram em situação de vulnerabilidade e necessitam de proteção. É necessário que exista transparência quanto aos riscos, possíveis benefícios, de modo que as expectativas estejam sempre claras, especialmente quando corporações, que buscam o lucro, estão envolvidas. O indivíduo fornecedor de dados, nem sempre é aquele que se beneficia dos mesmos. Os artigos aqui analisados apontam que pode ser inapropriado esperar relações de confiança ou profissionalismo entre sujeitos e corporações processadoras de dados, uma vez que a relação visa vantagens comerciais. Exemplos dessa preocupação são expressos nos seguintes trechos encontrados na literatura:

So how do we ensure, at an international level, that data sharing which is beneficial to the individual and to society can occur, in a way in which the rights of the individual (and/or the data producer) are appropriately addressed and the legal and ethical responsibilities are clearly delineated? (53, p. 67).

... as data-intensive research initiatives build on the use of data that patients provide not least with the intent to improve their own health, it is important to become more transparent about whether patients participate for their own medical benefit or whether participatory notions are evoked in the sense of public engagement that goes beyond individual purposes (52, p. 7).

D. Injustiça e desigualdades

O uso de Big Data envolve muitos custos de aquisição e tratamento de dados, além do custo operacional de computadores capazes de processar todas as análises. Essa é uma tecnologia cara:

Additionally, very high-throughput sequencing machines are very expensive (more than 1 million euros) and depreciation cannot be done over 2-3 years, which is a challenge for the economic model (63, p. 8).

O investimento necessário é um fator limitante e pode excluir a vasta maioria dos interessados em pesquisar ou trabalhar com isso. Consequentemente, observa-se um efeito concentrador na mão de algumas grandes corporações ou governos. Desse modo, a lacuna digital já existente pode se tornar ainda maior com a falta dos ganhos provenientes do uso de dados. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

... the term 'Big Data divide' is used to describe (...) the inequalities between data subjects providing the material for Big Data analytics and the organizations with the necessary infrastructure and resources to analyze and understand the data (1, p. 322).

Adicionalmente, em função da elevada demanda por investimento, pode ser que populações em condição de desvantagem financeira e social tenham acesso aos seus benefícios, o que por sua vez, gera um efeito de retroalimentação. A desigualdade entre estes e os que podem usufruir das suas vantagens é tendencialmente crescente:

... it must be noted that global health disparities could be magnified because of the digital divide among inter and intra nationally disadvantaged populations, which already carry a disproportionate burden of disease (55, p. 1012).

E. Modulação comportamental

Com a quantidade de dados coletada atualmente, a concentração de poder associada à capacidade de vigilância é um problema crescente, especialmente quando isto permite a alteração comportamental. A modulação de comportamentos se caracteriza como a utilização de dados com o intento de induzir e moldar o comportamento dos indivíduos dos quais se obteve as informações. Algoritmos são utilizados para reforçar visões de mundo, padrões de relacionamento e comportamento. O nexos entre a modulação do comportamento e a concentração de poder político e econômico atinge escalas planetárias e ameaça os regimes democráticos. De forma agravada, comumente, gera-se uma "via de informação", ou seja, o processo de vigilância não se limita a coletar a informação disponível, mas cria novos caminhos que induzem o fornecimento de mais dados, gerando uma retroalimentação contínua do poder de manipulação dos indivíduos. Exemplos dessas preocupações são expressos nos seguintes trechos encontrados na literatura:

Second, we need better guidelines around the ethical use of data, especially around profiling and social manipulation (64, p. 278).

... machine learning and other technologies that aim, in part, to forecast clinical and ecological outcomes can also result in social control or concentration of power and pan surveillance in planetary society (65, p 513).

F. Preconceito, estigma e injustiça

Uma vez que os dados são utilizados para classificar e prever comportamentos de inteiros segmentos da população, o estigma e a discriminação emergem como preocupação ética no *corpus* analisado:

Problematic discrimination and stigmatization of affected groups is therefore a real risk, even in anonymized datasets (1, p. 318).

While social stigmatization and discrimination can always result if medical data fall into wrong hands, the large-scale analysis of data from different sources increases the amount of sensitive information, e.g., by facilitating more fine-grained stratification (52, p. 2).

Um dos exemplos mais citados são possíveis práticas discriminatórias provenientes de seguradoras e no ambiente de trabalho:

Particularly when talking about sensitive data such as health information, people can worry about discrimination if it were available to insurance companies or employers (55, p. 1008).

Além disso, caso exista algum viés nos dados ou algoritmos, estes podem levar a perpetuação de abusos e estigmatizações cometidos contra grupos vulneráveis. Ainda com relação a minorias, pode ser que estas sofram mais à medida que se ampliem as rotinas de coleta de dados e suas aplicações. A própria situação de vulnerabilidade pode diminuir a capacidade dos sujeitos de controlar os dados que são coletados sobre si. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

These 'blind spots' mean that analysis will tend to focus on data subjects and phenomena amenable to digitization and measurement, meaning that the benefits and ethical burdens of Big Data will be placed, for better or worse, on specific social, cultural and economic groups (1, p. 325).

Os próprios algoritmos também são objeto de preocupação, já que podem possuir vieses recebidos de quem os desenvolveu ou não possuírem as devidas

medidas para evitar danos às análises. O viés algorítmico ou na coleta dos dados apresenta grande possibilidade de danos discriminatórios, tendo em vista que as injustiças presentes na sociedade possivelmente serão ampliadas e reforçadas:

Concerns for justice are raised in Big Data because prejudices or biases can be built in during the collection and coding of data (55, p. 1009).

The very nature of machine learning algorithms makes it plain that one might unintentionally develop a biased system or accept biased results (66, p. 27).

G. Tomada de decisão baseada em dados e responsabilidade moral

Tecnologias preditivas, quando alimentadas por grandes quantidades de dados, podem apresentar alta precisão. É justamente por este motivo que as tecnologias de Big Data são crescentemente utilizadas como auxiliares ou até mesmo soberanas nos processos de tomada de decisão. Do ponto de vista moral, no entanto, a autonomia de algoritmos e inteligências artificiais apresenta um desafio. A capacidade de contextualização limitada, a ausência da compreensão de que não existem absolutos morais e a dificuldade na responsabilização por resultados negativos são alguns dos limitantes na tomada de decisão por máquinas. O trecho a seguir exemplifica parte deste argumento no *corpus*:

Although there is growing evidence that AI may enhance predictive and diagnostic accuracy, it is uncertain whether intelligent software should be seen as an assistive tool in the hands of researchers and physicians or whether it should be rather attributed an autonomous decision-making role (61, p. 83).

Além da dificuldade em identificar o responsável por eventuais erros, a auditabilidade do processo da tomada de decisão também é um fator a se considerar. Esse é um aspecto inclusive importante na reflexão sobre a regulação ética em pesquisa. Isso porque, enquanto as pesquisas realizadas com seres humanos passam por crivos éticos, as conclusões tiradas de pesquisas com dados secundários não passam normalmente por esse tipo de revisão. O trecho a seguir exemplifica o problema:

Specifically, one issue central to the discussion of privacy ethics and policy is that of accountability. While biomedical research for academic purposes is subject to institutional review, uses of information gleaned from secondary data mining processes often exist outside the jurisdiction of conventional ethics boards (62, p. 304).

Outro desafio crucial reside no propósito comercial da tecnologia. As corporações que atuam na coleta e armazenamento dos dados via de regra o fazem com o propósito de lucro e não por motivações benevolentes ou para promover o bem maior da sociedade. Dessa maneira, a tomada de decisões baseada nesses dados pode acabar por refletir outros interesses que não o interesse público:

The other issue on which there was disagreement was whether it was the role of the data custodian to assess the public interest (67, p. 508).

H. Validade do conhecimento

Levando em consideração todos esses aspectos levantados até o momento, fica evidente que se não realizadas com todo o rigor necessário, pesquisas clínicas e científicas podem perder sua validade. Um exemplo bastante significativo disso é que, se o uso de tecnologias de aprendizado de máquina em pesquisa clínica torna seus percursos metodológicos opacos aos critérios científicos de reprodutibilidade, o conhecimento produzido tem problemas quanto à validação:

Of note, the most used AI in rheumatology was machine learning (ML) methods, although most papers did not state whether the algorithms were supervised or not. This is a first concern in the domain of ethics, since complete reports of research methods are necessary for reproducibility (55, p. 1010).

A origem e confiabilidade dos dados também é um problema para a validade do conhecimento produzido. O uso de dados secundários em pesquisa pode negligenciar a observância de aspectos éticos e rigor técnico e científico na sua extração original:

Improbable claims about direct, automated data transfers from confidential hospital EHRs to supply 'big' proprietary health databases are increasingly appearing in the grey literature. These need to be tempered with the realities of confidential patient information and its management, and the actual data items held (68, p. 4).

O foco no respeito à privacidade, promove desmobilização e negligência quanto a outros fatores éticos importantes. Um exemplo dessa preocupação é expresso no seguinte trecho encontrado na literatura:

Yet, in discussions about information-based research, I think there is often too great a focus on privacy as the sole criterion of ethical research (69, p. 427).

5.4.3 Classe 3 – Dimensões Econômicas

Nas dimensões econômicas, foram identificados como aspectos éticos no *corpus* analisado: comoditização da privacidade; e percepção pública.

A. Comoditização da privacidade

Um tema que emerge com o uso de Big Data e a comercialização de agregados de dados e de tecnologias preditivas é aquilo que se tem chamado de *comoditização da privacidade* (70).

O fenômeno se caracteriza pela crescente comercialização de dados obtidos por entes privados. Nesses casos ocorre um desvio da finalidade, uma quebra na confiança e se frequentemente uma violação ao consentimento de quem forneceu o dado sem ter plena consciência de que ele seria usado para outros fins. No campo da saúde, isso é especialmente sensível, uma vez que o uso de dados anonimizados obtidos a partir de prontuários e outros registros não é suficiente para evitar esse problema.

Os dados de saúde privados e compartilhados em nome de uma relação de confiança com profissionais de saúde passam a ser comercializados. A prática gera valor econômico sem consentimento e corrompe a motivação coletiva das práticas de saúde pública, como pode ser apreendido nos seguintes trechos:

When Big Data is thought of as a form of business based around the selling and processing of data for commercial advantage, it is perhaps inappropriate to expect a relationship based on 'trust' or professionalism to exist between subjects and custodians (1, p. 327).

This may raise concern among patients about industry access and commercialization (55, p. 1009).

The commercial use of health data can be divided into a tripartite taxonomy of concerns: (1) *data misuse*, such as privacy intrusion, increased cost through price stratification, and economic or employment discrimination; (2) *profiteering*, *i.e.*, exclusive private benefit without any public benefit; and (3) *commercialization*, *i.e.*, corruption of the public service ethos of an organization such as the NHS (71, p. 13).

B. Percepção pública

Dada a grande participação de grandes corporações de tecnologia, que possuem lucros cada vez maiores, o público pode não consentir em partilhar seus dados, o que, por sua vez, pode significar um atraso para a ciência como um todo:

Particularly when talking about sensitive data such as health information, people can worry about discrimination if it were available to insurance companies or employers (55, p. 1008).

Na literatura, o aspecto ético “percepção pública” implica na percepção de praticamente um consenso quanto aos benefícios dessas tecnologias, e ao risco que usos moralmente duvidosos possam influenciar negativamente a percepção pública:

... the erosive use of Big Data for profit in business and industry has polluted public perceptions of all uses of Big Data (66, p. 28).

Desse modo, a manutenção da confiança do público é cada vez mais desafiadora, levando a piores coletas de dados e menor aceitação de resultados pela sociedade:

As studies show, for fear of disclosure to unauthorized persons, patients may withhold information, giving an incomplete or misleading description of their condition (59, p. 93).

É de se destacar que essa preocupação, conforme observada nos artigos analisados, parece mais interessada em evitar uma não adesão à tecnologia, do que uma participação e debate coletivo sobre ela.

5.4.4 Classe 4 – Dimensões Regulatórias

Nas dimensões regulatórias, foram identificados como aspectos éticos no *corpus* analisado: consentimento; mecanismos de proteção; e privacidade e confidencialidade.

A. Consentimento

Por pressupor uso massivo e entrecruzado de dados de diversas fontes, as tecnologias de Big Data configuram um desafio para o consentimento informado em sua maneira tradicional, tendo em vista que é virtualmente impossível prever todos os usos futuros dos dados. Alternativas vêm sendo estudadas para se adequar a esta tecnologia, mas ainda constitui um dos principais desafios de seu uso. É válido frisar a inadequação de alguns métodos para obtenção de consentimento atualmente, especialmente quando feitos por meio de termos de serviços e contratos de usos de aplicativos ou serviços. Outro desafio são coletores de dados que solicitam o consentimento, mas o fazem de forma não clara para não comprometer a segurança das instituições que representam. Exemplos dessa preocupação são expressos nos seguintes trechos encontrados na literatura:

This form of consent is problematic because Big Data is intended by design to reveal unforeseen connections between data points. This means that both what the data reveals about the subject and its utility in future research present greater uncertainty than normal at the time of consent (1, p. 312).

... it is not clear whether informed consent can truly be obtained if the patient cannot be given a full explanation of how a machine is used (55, p. 1010).

... single instance use consent is difficult (72, p. 592).

Via de regra, o consentimento dos indivíduos que participam em pesquisa é um requisito ético da mesma, tendo em vista possíveis riscos e danos. Um grande desafio quanto à coleta de dados para Big Data é que muito do “consentimento” obtido por quem realiza as coletas não é explícito, podendo estar oculto no meio de termos de serviços, ou mesmo inexistente:

It is apparent that terms of service agreements, click-through consents, and voluntary oversight are inadequate (69, p. 426).

Quando a coleta ocorre de forma passiva, sem que o indivíduo perceba, este é usurpado quanto a sua escolha de quais informações deseja partilhar. Ainda mais desafiador é o fato de a grande maioria da sociedade não estar ciente de quais dados são coletados sobre si mesmos e que usos são dados a eles, como exemplificado pelos trechos a seguir:

Overall, the ethical challenge here is to ensure valid consent models preventing the use of data or biomaterials without the knowledge or consent of participants or even against their explicit will (52, p. 5).

With regards to comprehension, most users are not aware of the multiplicity algorithms that are gathered, stored and analyzed from their data by multiple agents (59, p. 101).

... some research using social media has been criticized for a lack of informed consent ... (69, p. 425).

B. Mecanismos de proteção

Como diferencial, as tecnologias de Big Data oferecem inúmeras possibilidades de resultados com um mesmo conjunto de dados, o que por si só dificulta significativamente uma regulação específica para controlar seu uso. Com isso, *frameworks* e procedimentos éticos utilizados anteriormente podem estar desatualizados e despreparados para lidar com as realidades de Big Data:

The possibility given by technology to link various data from pathogen samples challenges the existing legal and ethical principles (73, p. 693).

Do mesmo modo, a legislação nem sempre está ajustada a essa nova realidade, podendo possuir até mesmo regulações inconsistentes em diferentes níveis de governo:

From a regulatory perspective, further challenges arise due to inconsistent regulations or interpretation thereof across federal agencies and offices governing research (74, p. 3).

No meio privado, a situação é ainda mais grave. A regulação de atividades econômicas é, em geral, menor do que aquelas governamentais. Assim, entidades que visam o lucro, sujeitas a pouca ou nenhuma ação regulatória externa a depender do país, atuam como agentes auto reguladores. A dificuldade em saber inclusive quais dados são coletados, por quanto tempo, para quem são vendidos e quais são seus usos tornam os mecanismos de proteção virtualmente obsoletos. É atualmente um desafio definir limites para este tipo de atividade e igualmente difícil propor maneiras de garantir que sejam cumpridos, como abordado nos trechos a seguir:

Ethical frameworks and procedures developed half a century ago for controlled experiments in universities and hospitals struggle when faced with

real-time analysis, productization, and monetization of the incalculable “data exhaust” we produce each day with our digital devices (75, p. 4).

It seems that many patients are unaware of possible conflicts of interest regarding data sharing, including for commercial purposes, and existing protective laws and ethical arguments do not fully address these new challenges (76, p. 2).

A coleta de dados, inclusive de saúde, em redes sociais é um agravante na dificuldade de regulação, uma vez que enquanto as corporações agem digital e transnacionalmente, os Estados são limitados por suas jurisdições. Exemplos desse desafio são expostos nos trechos que seguem:

Of immediate concern is that the use of personal information linked to health or, even worse, the intentional manipulation of behavior, is not subject to traditional, federal research oversight (69, p. 425).

Next, insist on accountable oversight. From bioinformatics laboratories to hospitals, those who collect, store, and analyze health information are stewards of a precious resource. They must, absolutely must, be held to trust-enhancing standards (66, p. 29).

This raises a number of ethical concerns, including a lack of ethical oversight of data use in private corporations... (54, p. 273).

Mesmo quando os dados são coletados por entes públicos, nem sempre existe a transparência quanto à finalidade, ao atual uso dos dados, e ao possível destino após o uso das informações:

To what extent should public health entities promote data transparency, share data with other entities (e.g., commercial, law enforcement), or provide mechanisms of control for those represented in the data? (77, p. 537).

We need to demonstrate integrity, honesty, and transparency as to what happens to data and what level of control people can, or cannot, expect (78, p. 1014).

Third, the use of AI by public bodies brings calls for transparency (79, p. 165).

É bastante significativo na literatura associada à regulamentação de Big Data o debate em torno da propriedade dos dados. Muitos indivíduos possuem o forte sentimento que seus espécimes biológicos e dados são seus, de modo que deve ser claro como funcionará a relação entre quem fornece, utiliza e (sendo o caso) lucra com isso:

Understood in terms of 'control', ownership grounds empowerment of data subjects through mechanisms to track and check the existence and manipulation of their data, which can help prevent "the existence of "secret" databases and leverage societal pressure to constrain any unacceptable uses" (1, p. 319).

Contudo, ainda que o conceito de propriedade seja claramente aplicável a bens materiais, quando transportado para o meio digital, identificar a quem pertencem os dados e garantir que essa posse seja respeitada é bastante desafiador. Mesmo definindo que a propriedade é de determinada entidade ou indivíduo, ainda não é claro o que isso representa para a coleta e análise de dados.

C. Privacidade e confidencialidade

A violação da confidencialidade e privacidade implica em risco de outros danos. Assim, as ameaças a privacidade de um indivíduo devem ser mitigadas para que dados pessoais não caiam em mãos erradas:

The commercial use of health data can be divided into a tripartite taxonomy of concerns: (1) data misuse, such as privacy intrusion... (71, p. 13).

The major risk to patients is the exposure of their personal health information (80, p. 105).

Talvez o risco mais frequentemente mencionado por diversos autores é o risco de reidentificação após a coleta de dados. Nos primórdios de Big Data era comum que apenas técnicas de anonimização fossem empregadas e que estas fossem o suficiente. Atualmente, como os bancos de dados estão cada vez mais interconectados, aumenta o risco de reidentificação, o que atinge diversamente indivíduos e grupos vulneráveis, conforme apresentado nos trechos a seguir:

When the possibilities of re-identification and 'hidden' analysis exist, data subjects' control over uses of their data acquires greater importance, as control allows subjects to restrict undesired uses (1, p. 319).

Current privacy rules do not deal with longitudinal data and transactional data, which can be used to re-identify an individual (62, p. 296).

Data linkage exacerbates some of the ethical concerns related to privacy and confidentiality, since the collection of increasingly detailed information about patients, from more varied sources, may also increase the risk of re-identification — matching anonymized data with other information to uncover the identity of their owners (55, p. 1009).

6 DISCUSSÃO

A amostra da literatura científica selecionada e o Relatório do IBC ofereceram elementos suficientes para compreender o presente estado sobre a reflexão acadêmica a respeito das implicações éticas do uso de Big Data em saúde. Embora o Relatório do IBC e os artigos tenham algumas diferenças quanto à profundidade da abordagem, e existam aspectos mutuamente ausentes quando comparados, a compreensão dos principais aspectos da dimensão ética do fenômeno Big Data pareceu bastante convergente. Metodologicamente, o processo de análise foi facilitado pela utilização da ferramenta auxiliar, o IRaMuTeQ. A metodologia delineada ofereceu meios para perseguir o objetivo geral desta pesquisa, que foi o levantamento das principais implicações éticas da utilização de Big Data no campo da saúde humana já identificadas pela atual produção científica sobre o tema. O objetivo específico de identificar e comparar os desafios também foi facilitado pelo percurso metodológico eleito, que permitiu uma categorização mais evidente deles.

O Relatório do IBC sobre Big Data e Saúde (4) é, como era de se esperar pela *expertise* dos membros do conselho, um documento amplo que de fato antecipa quase todas as principais implicações éticas da utilização de Big Data em saúde. Como visto nos resultados, o IBC identificou a Autonomia e consentimento; Privacidade e confidencialidade; Da propriedade à custódia e compartilhamento de benefícios; Justiça e a lacuna digital; Justiça e a não discriminação; Sustentabilidade em relação à energia e ao meio ambiente; e Pesquisa, como os principais desafios na utilização de Big Data em saúde.

Observou-se que as temáticas (e conseqüentemente os desafios presentes) identificadas pelo IBC, vão de questões facilmente perceptíveis (como as ameaças à privacidade), até pontos sequer considerados pela literatura encontrada (como o impacto ambiental do uso dessa tecnologia). Deste modo, a leitura do relatório apresenta-se como uma enriquecida introdução à reflexão ética sobre o tema.

Entretanto, conforme o próprio Relatório afirma em seu início, o texto não pretende ser exaustivo. Ressalta-se que dois dos desafios mais significativos, principalmente do ponto de vista de uma bioética produzida em um país periférico e preocupada com a desigualdade, foram tratados apenas na literatura analisada. A comoditização da privacidade e a modulação do comportamento, aspectos com

profundas implicações para a questão da justiça, não figuram no Relatório. Dessa maneira, embora o documento expresse continuamente a preocupação com a injustiça, a discriminação, a dificuldade de apropriação do conhecimento em contextos periféricos, a necessidade de compartilhamentos de benefícios, o mecanismo perpetrador de injustiça nessas tecnologias passou ao largo da análise do Relatório.

É possível aventar ao menos duas razões para que o documento não mencionasse tais aspectos. A primeira delas, é que esse posicionamento pode ter sido resultado de uma necessária negociação em contextos multirepresentativos, como de fato é o Comitê. Outra explicação pode ser a própria temporalidade da discussão. A autora mais reconhecida que pauta o tema da comoditização da privacidade, Zuboff (70), ganha notoriedade com a publicação de seu livro *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* em 2019, posterior ao relatório do IBC, portanto.

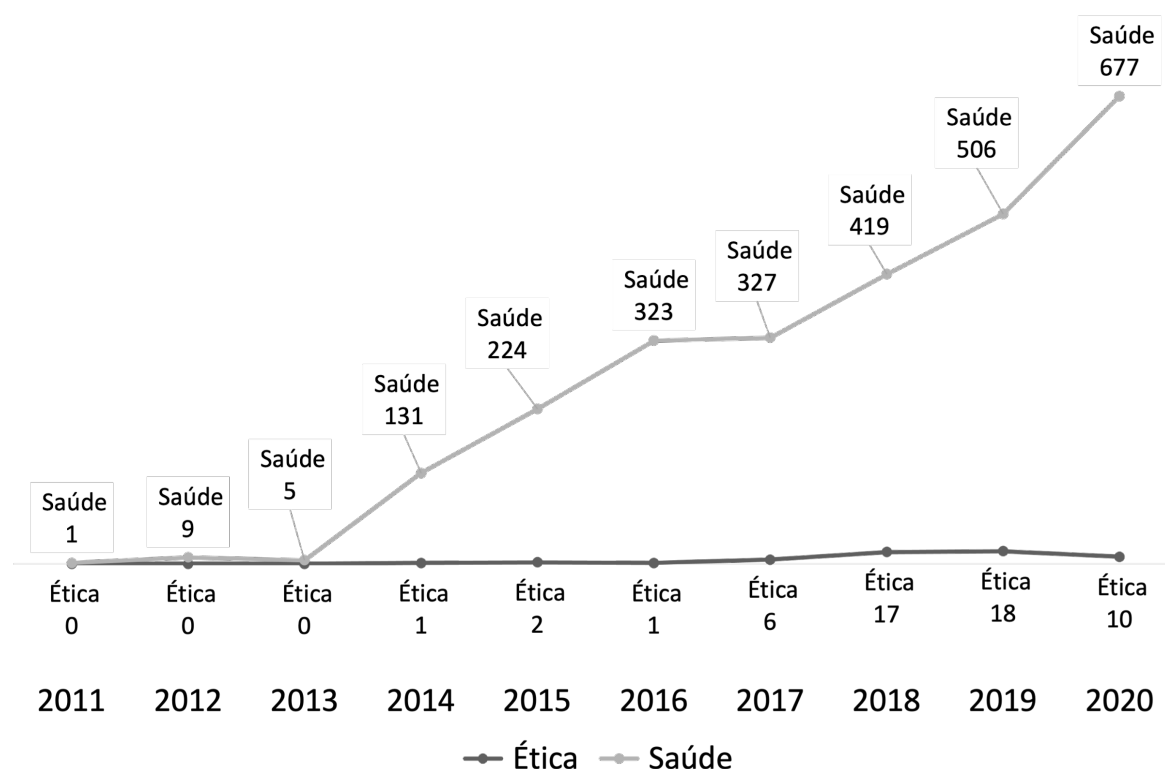
A discussão, iniciada pela autora estadunidense, teve uma recepção bastante significativa e representa uma crítica dura à nova face do capitalismo neoliberal caracterizada justamente pelo processo de comoditização da privacidade. Adiante, a presente discussão tentará delinear alguns eixos a partir dos quais futuras leituras podem aprofundar as maneiras como esse novo contexto pode afetar os modos de vida nos países e contextos periféricos. A compreensão sobre o tema parece bastante significativa, uma vez que as tecnologias de Big Data, em contínua expansão, terão cada vez mais importância no campo da saúde, com implicações bioéticas cada vez mais importantes.

6.1 CONTEXTUALIZAÇÃO DO ESTADO ATUAL DA LITERATURA

O debate sobre a moralidade do uso de Big Data em saúde ainda é muito incipiente. Sinal inequívoco é que o volume de publicação científica sobre a dimensão ética do uso de Big Data em saúde está muito aquém do incremento nas publicações técnicas sobre o tema nos últimos anos. Ao se comparar os artigos obtidos como resultado da busca dos descritores “Big Data” e “Saúde”, nas mesmas bases até o ano de 2020, foram encontrados 2615 artigos na base PubMed (contra os 55 artigos desta base durante a pesquisa) e sete artigos na base LILACS (contra a ausência de

artigos nesta base durante a pesquisa). Com isso, em face das significativas implicações éticas sobre o tema, o debate ético certamente precisa se intensificar. Abaixo, na Figura 10, podemos comparar a tendência quanto à produção técnica e a discussão ética sobre o uso de Big Data em saúde na literatura científica.

Figura 10 – Comparação da tendência entre a produção técnica e a discussão ética sobre o uso de Big Data em saúde.



Fonte: Gráfico elaborado pelo autor com dados obtidos a partir da busca dos descritores “Big Data” e “Saúde” nas bases científicas PubMed, LILACS e dos artigos analisados nesta dissertação.

Ressalta-se que, mesmo de pequena expressão nas bases pesquisadas, a preocupação ética apresenta tendência crescente, conforme mostra o incremento de artigos desde 2015, quando aparecem os primeiros dois artigos analisados no *corpus*. A tendência permanece a mesma até 2020, quando, em comparação ao ano anterior, ocorreu uma significativa queda no número de publicações. Uma das possíveis explicações para tal fenômeno é alguma demora que sempre precede a indexação dos artigos por parte das bases. No entanto, pode-se conjecturar que tenha sido em

parte resultante das consequências inéditas da pandemia de COVID-19 sobre a dinâmica global da pesquisa e publicação científica (81).

Em relação à nacionalidade da primeira autoria dos artigos, conforme apresentado na seção de Resultados, observou-se uma produção principalmente anglo-saxônica com predominância estadunidense, com 19 (43,18%) dos 44 artigos selecionados. Esse padrão segue, de maneira geral, fenômeno observado globalmente na editoria e publicação científica, ou seja, que os países centrais lideram com larga vantagem o mundo acadêmico, com destaque para os Estados Unidos, que possuem 21,36% da publicação científica mundial (82). Tal dinâmica acaba por privilegiar um enfoque em discussões que refletem preocupações, posicionamentos e desafios éticos centrados nos contextos do norte geopolítico. Isso é ainda mais significativo no caso do Big Data, uma vez que esse é um tipo de tecnologia que demanda altos recursos técnicos e econômicos e seus principais polos desenvolvedores são sediados no Vale do Silício.

Garrafa (83) sinaliza, no entanto, para a necessidade de que a bioética no Brasil – e na América Latina como um todo – produza reflexões locais e se recuse a “importar” de forma acrítica e sem contexto as preocupações éticas estrangeiras. Uma das características resultantes desse posicionamento foi a politização e ampliação de escopo das bioéticas latino-americanas. Desde os anos 90, a bioética latino-americana já trabalhava a ampliação de sua agenda, tendo papel fundamental na conseqüente transformação pela qual tem passado a bioética internacional. Esse processo resultou na DUBDH nos moldes que conhecemos hoje (84).

No caso da reflexão acerca do uso de Big Data em saúde, ainda há muito trabalho a se fazer nesse sentido. A incipiência é ainda mais significativa em bases regionais e nacionais. É significativo o fato de que, observando os critérios de busca, não foram encontradas publicações regionais e nacionais sobre o tema. Evidentemente, o resultado não permite a extrapolação de uma completa inexistência de discussão sobre o tema na América Latina. O que o dado reflete, sem dúvida, é que nos contextos nacional e latino-americano a incipiência na discussão ética sobre o uso de Big Data em saúde é ainda maior do que em nível global.

Conclusão semelhante foi obtida por Harayama (85), que na busca com descritores “Big Data”, “Bioética” e “Ética”, de maio a julho de 2020 nas bases SciELO, BIREME e Jstor, bem como incluindo na pesquisa sítios da Organização Mundial da Saúde e do Ministério da Saúde do Brasil, encontrou somente 21 artigos (85). A

própria existência desse artigo apenas citado é prova de que existe reflexão e que o processo de indexação por bases científicas representa uma adicional dificuldade que se interpõe a um maior debate regional sobre o tema.

O *corpus* foi também analisado quanto à área de atuação científica dos seus primeiros autores, conforme apresentado na seção de Resultados. Nesse aspecto, identificaram-se cinco áreas distintas, destacando-se as áreas de saúde e tecnologia. Tal resultado não surpreende, uma vez que a busca se deu em bases da área da saúde e que a pesquisa estuda uma tecnologia em ascensão. Ainda, produções classificadas nas áreas de ética/bioética e direito foram encontradas na literatura.

Com o apoio do IRaMuTeQ, foram identificadas classes a partir da análise dos artigos encontrados na literatura. O agrupamento em quatro classes, resultante da categorização realizada pela aplicação, favoreceu a compreensão do caráter multifacetado das implicações éticas em análise.

Apesar do auxílio que o agrupamento em classes forneceu, uma limitação do software de apoio reside na necessidade de considerável conhecimento prévio sobre o tema a ser analisado. Isso ocorre uma vez que o software não fornece um relatório único sobre o *corpus* e sim uma série de dados direcionados (por exemplo, o nível de importância de uma palavra específica para uma classe, quais palavras mostraram maior predominância em uma classe e quais segmentos de texto encontram-se em uma classe). Os tipos de resultados fornecidos pelo software, sem o mencionado conhecimento prévio do tema em análise, possuem relevância limitada.

A compreensão de um fenômeno real, como é o conjunto de textos produzidos sobre ética e Big Data em saúde, não pode ser plenamente desenvolvida lançando mão apenas do software. Os argumentos, amplamente encontrados na literatura, de que o uso de Big Data demanda treinamento prévio e cuidados no processo de interpretação de resultados foram observados no próprio emprego da metodologia desta dissertação. As ferramentas de tratamento e análise de dados não oferecem "resultados prontos". A compreensão e interpretação dos resultados ainda são aportes humanos, em seus erros e acertos.

É válido notar que as classes de dimensões éticas identificadas na literatura, assim como os aspectos éticos entre si, guardam inter-relação. Isso faz com que aspectos éticos categorizados em outras classes possuam facetas econômicas, quer em suas causas quer em seus resultados. Exemplo disso, é o aspecto chamado "Injustiça e desigualdades", a questão, de cunho social, é comumente atribuída na

literatura aos elevados custos associados ao desenvolvimento da tecnologia. A seguir um dos trechos que ilustra o nexó entre a dimensão econômica e a social:

... it must be noted that global health disparities could be magnified because of the digital divide among inter and intra nationally disadvantaged populations, which already carry a disproportionate burden of disease (55, p. 1012).

Da mesma maneira, aspectos classificados como dimensões econômicas só o são porque seu efeito na coletividade pode gerar prejuízos econômicos. Exemplo disso, é o aspecto denominado “percepção pública”. A percepção pública é frequentemente mencionada na medida em que os artigos apontam para a necessidade de que essa seja positiva. Revela-se que a preocupação tem mais a ver com a promoção da adesão da tecnologia (com finalidades de consumo) do que com a real busca por uma reflexão por parte da sociedade acerca de suas consequências.

... the erosive use of Big Data for profit in business and industry has polluted public perceptions of all uses of Big Data (66, p. 28).

As classes assim foram apropriadas no estudo como orientadoras de uma compreensão mais organizada do fenômeno, não encaradas como divisões estanques. Essa aproximação entre elas foi inclusive evidenciada pelo gráfico de Análise Fatorial de Correspondência (AFC) que, conforme mencionado nos Resultados, é uma representação gráfica da associação das classes (e suas palavras) entre si.

No gráfico AFC, a maior aproximação semântica entre os discursos se deu entre as classes 2 (Dimensões sociais) e 3 (Dimensões econômicas). A associação entre as duas dimensões, de alguma forma esperada em qualquer fenômeno coletivo amplo, é incrementada no caso do uso de Big Data. Isso porque nesse advento tecnológico os produtos e serviços comercializados são justamente fruto da agregação massiva de dados de pessoas e grupos. Consequentemente, a relação entre a dimensão coletiva e econômica dessas tecnologias é bastante incrementada.

Já a Classe 1, Dimensões digitais, foi a que mais se distanciou das demais no gráfico, o que demonstra uma distinção mais pronunciada entre os discursos e preocupações éticas expostas. A primeira classe trata mais de perto de desafios técnicos, como a condição humana enquanto ser digital, incerteza e imprevisibilidade intrínsecas à tecnologia, cooperação para produção de informações, opacidade das

tecnologias e das corporações econômicas que as comercializam, segurança digital e qualidade e representatividade do conhecimento produzido. Essa classe, que concentra uma maior intimidade entre aspectos propriamente técnicos e o surgimento dos problemas éticos, teve predominância de autoria estadunidense. Esse foco em uma abordagem mais técnica por parte dos EUA pode ser ao menos parcialmente atribuída ao fato de ser esse o principal polo tecnológico do mundo sobre o tema.

Nas demais classes, observou-se a predominância de artigos provenientes de países europeus. Nesses países, a tradição regulatória referente à proteção dos direitos humanos culminou com a primeira legislação mundial recente que visa a proteção de dados, a Regulamentação Geral de Proteção de Dados da União Europeia (86). Essa preocupação pode explicar a predominância de países europeus quando se discutem as dimensões sociais, econômicas e regulatórias do uso de Big Data em saúde, encontrada neste estudo.

Ainda que limitado como amostra para extrapolar qualquer interpretação, o único artigo de autoria brasileira tratou de dimensões técnicas. O artigo representa um editorial, que aparece na base SciELO, mas foi oferecido como resultado apenas na busca da PubMed, e não na LILACS, a base latino-americana que incorpora a SciELO. Esse problema na indexação dos resultados de busca pode ser mais um fator que venha a prejudicar a discussão nacional e regional sobre o tema, como já mencionado.

Sobre o conteúdo, o texto de autoria brasileira, de grande apelo técnico, aponta como desafios do Big Data o processo de autorização para coleta de dados automáticos, procedimentos de encriptação para proteção de dados e o risco quanto à validade das associações encontradas com o teste automático de hipóteses aleatórias na área da saúde (87). A modesta produção nacional e o enfoque em aspectos técnicos e não em sociais ou regulatórios parece em descompasso com nossa realidade e o tipo de preocupação ética que o uso dessas tecnologias em contextos periféricos demanda.

Tendo em consideração as tendências mais amplas dos resultados acima, passa-se neste ponto à análise comparativa dos aspectos éticos apontados pelo Relatório do IBC sobre Big Data e Saúde (4) e aqueles identificados na literatura científica analisada.

6.2 ANÁLISE COMPARATIVA

Na Tabela 3, é apresentada uma síntese da correspondência entre os desafios listados no Relatório do IBC sobre Big Data e Saúde (4), e os resultados da análise da literatura.

Tabela 3 – Correspondência entre os desafios do Relatório do IBC sobre Big Data e Saúde (4), as dimensões atingidas por estes e os respectivos aspectos éticos identificados na literatura.

Desafio do IBC	Dimensões atingidas	Aspecto ético
A. Autonomia e consentimento	Digital	Incerteza e imprevisibilidade
	Social	Apropriação do conhecimento
		Tomada de decisão baseada em dados e responsabilidade moral
	Regulatória	Consentimento
B. Privacidade e confidencialidade	Digital	Opacidade
		Segurança digital
	Econômica	Percepção pública
	Regulatória	Privacidade e confidencialidade
C. Da propriedade à custódia e compartilhamento de benefícios	Digital	Cooperação
	Social	Conflito entre interesses individuais e coletivos
	Regulatória	Mecanismos de proteção
D. Justiça e a lacuna digital	Social	Injustiça e desigualdades
E. Justiça e a não discriminação	Digital	Qualidade e representatividade
	Social	Criação de perfis e autonomia Preconceito, estigma e injustiça
F. Sustentabilidade em relação à energia e ao meio ambiente	-	-
G. Pesquisa	Social	Validade do conhecimento

Fonte: Esquematização realizada pelo autor a partir dos tópicos de desafios, dimensões que constam na literatura encontrada e no Relatório do IBC sobre Big Data e Saúde

Tanto os aspectos relatados pelo IBC quanto os identificados na literatura científica foram descritos, de forma independente entre si, na seção de Resultados.

A análise do relatório e do *corpus* leva à compreensão de que as implicações apontadas explicitamente pelo Relatório, por vezes, circunscrevem mais de um aspecto especificado na literatura. Essas relações e intersecções foram sumarizadas na Tabela 3 acima e são transcritas sucintamente abaixo.

É possível compreender o tratamento realizado pela ferramenta acessória de análise como uma desagregação das dimensões éticas específicas em que as preocupações se manifestam.

Em resumo, quando em análise comparativa chegou-se à compreensão expressas nos seguintes parágrafos.

- A. Quanto à autonomia e consentimento, poderíamos afirmar que, devido às dimensões digitais, a incerteza e imprevisibilidade intrínseca ao caráter expansivo das tecnologias digitais apresentam-se como desafios à autonomia dos sujeitos, impossibilitados de decidir de forma plenamente ciente. Do ponto de vista social, a literatura apontou que a capacidade de apropriação do conhecimento também impacta a autonomia dos indivíduos e pode ter efeitos de desigualdade de acesso. Importa também no caráter coletivo a discussão acerca da autonomia como um atributo a ser reconhecido nas decisões das máquinas, o que pode gerar dificuldade na compreensão de quem é responsável moral pelas escolhas. Todos esses aspectos refletem, do ponto de vista regulatório, uma preocupação com o consentimento, suas formas de obtenção e validade, diante de tecnologias tão complexas e de usos virtualmente imprevisíveis.

- B. Já o desafio “Privacidade e confidencialidade”, identificado no Relatório do IBC, apareceu na literatura em quatro desafios distintos: “Opacidade” e “Segurança digital”, de natureza digital; “Percepção pública”, de natureza econômica; e, “Privacidade e confidencialidade”, de natureza regulatória. Do ponto de vista digital, a opacidade implica não compreender plenamente os processos que levam aos resultados quando se utiliza aprendizado de máquina. Da mesma forma, não há clareza quanto aos dados pessoais que as corporações detêm. Dessa maneira, em contraste à falta de privacidade que resulta em uma transparência imposta aos usuários, há uma opacidade quase completa quanto às corporações que usam e comercializam os dados. Do ponto de vista técnico, a segurança digital também é uma preocupação, uma vez que os acessos indevidos a dados parecem ser a origem da maior

parte dos problemas com o uso de Big Data. Do ponto de vista econômico, observa-se na literatura que a percepção pública é capaz de impactar o rumo de projetos envolvendo Big Data, uma vez que a queda da confiança do público pode impedir o compartilhamento de dados em projetos futuros. Esses aspectos refletem, do ponto de vista regulatório, em uma preocupação com a proteção da privacidade do indivíduo.

- C. Identificado como um desafio específico, o tema “Da propriedade à custódia e compartilhamento de benefícios” apareceu nos artigos analisados dissociado em três desafios: “Cooperação”, de dimensão digital; “Conflito entre interesses individuais e coletivos”, de natureza social; e, “Mecanismos de proteção”, de natureza regulatória. Em sua dimensão digital, a cooperação é abordada pela literatura ao discutir o acesso a dados. Os meios técnicos devem prover formas cuidadosas para que o compartilhamento de dados se viabilize. Os conflitos entre interesses individuais e coletivos são analisados dentro das dimensões sociais, uma vez que um correto equilíbrio entre direitos individuais e coletivos é fundamental para o compartilhamento real dos benefícios oriundos do uso de Big Data. Ainda, para que isso ocorra, é necessário que mecanismos de proteção, objetos analisados do ponto de vista regulatório, existam e estejam sempre atualizados. A atualização e adequação constante dos instrumentos regulatórios é um aspecto bastante citado na literatura.
- D. O desafio do IBC “Justiça e a lacuna digital” também aparece na literatura, e recebeu no estudo o título “Injustiça e desigualdades”, de natureza social. O uso de Big Data pode levar a um aumento na desigualdade já que existem grupos de indivíduos que apenas fornecem seus dados aos detentores das plataformas de coleta sem receber compensações proporcionais. Há, como resultante, um acúmulo de capital intelectual e econômico.
- E. O relatório do IBC torna a tratar da justiça nas aplicações de Big Data em outro desafio, “Justiça e a não discriminação”. Na literatura, tal

conteúdo aparece mais detalhado e subdividido em três aspectos específicos: “Qualidade e representatividade”, de natureza digital; “Criação de perfis e autonomia” e “Preconceito, estigma e injustiça”, de natureza social. Do ponto de vista digital, a discriminação é percebida como preocupação na literatura quando esta trata da falta de representatividade na coleta dos dados, que pode perpetuar vieses e preconceitos dos desenvolvedores. Ainda, do ponto de vista social, a justiça e a não discriminação são trabalhadas pela literatura ao comentar a criação de perfis, ou seja, a criação de “*personas*” para a análise não apenas de um indivíduo, mas de um grupo de indivíduos, atribuindo características comportamentais baseadas em estereótipos. Por fim, a divulgação de dados pessoais, como os de saúde, por exemplo, podem contribuir com estigmas e aprofundamento de injustiças.

- F. Um aspecto a se destacar é que um dos desafios identificados no IBC, “Sustentabilidade em relação à energia e ao meio ambiente”, não foi captado na análise da literatura. É interessante ressaltar que a preocupação ambiental alude à DUBDH, e um conjunto ampliado de preocupações bioéticas. Esse movimento de expansão, pelo qual tem passado a disciplina e que culmina na própria Declaração, possui grande contribuição da bioética latino-americana, em especial da Bioética de Intervenção, que desde seus primórdios busca propor uma leitura mais engajada dos conflitos bioéticos. A Bioética de Intervenção busca justamente repolitizar os conflitos morais e posicionar-se do lado dos mais vulneráveis (17) ao passo que sai do alcance apenas da clínica médica e chega ao mundo da vida em geral, recuperando de maneira mais problematizada as preocupações bioéticas com o meio ambiente que participaram da fundação da própria disciplina (88).
- G. Outro aspecto de divergência é a compreensão da relação entre o conhecimento científico e a ética. O Relatório do IBC aponta uma preocupação com a necessária promoção de pesquisas científicas, ressaltando para isso que a proteção de dados sensíveis pode se apresentar como obstáculo ou ponto de conflito. Já a literatura apresenta

abordagens mais críticas, apontando que por vezes a validade do conhecimento produzido é questionável. Devido à opacidade das análises, os critérios de reprodutibilidade e falseabilidade, necessários ao empreendimento científico, ficam impossibilitados, o que pouca dúvida sobre a validade dos resultados.

O que se depreende desse exercício de traçar relações entre o Relatório e a literatura é que, ainda que o Relatório do IBC apresente como um diferencial a preocupação ambiental, não detectada na literatura, a quantidade de aspectos éticos identificados nesta é mais elevada. Tal resultante é de alguma forma esperado. O *corpus* de artigos científicos analisados é um conjunto maior de textos, com diversidade de interesses por parte de autores de origens e campos acadêmicos diversos. É esperado que esses autores, em conjunto, tenham identificado uma maior quantidade de desafios, e até mesmo aprofundado as abordagens dos aspectos, já que não precisavam doar o aspecto abrangente e introdutório que o Relatório precisava ter. O intervalo no tempo contribui também para isso, uma vez que após o Relatório, a discussão na literatura, influenciada até mesmo pela publicação do próprio documento, pode ter ganhado mais fôlego.

O Relatório do IBC já adverte em sua apresentação que não pretende ser exaustivo quanto às possíveis implicações éticas do uso de Big Data em saúde.

Contudo, mesmo relevando sua intenção mais introdutória ao tema, o Relatório poderia ter abordado de maneira mais clara as divergências de posições quanto aos aspectos éticos do uso de Big Data em saúde que identificou. Por vezes, o relatório, com a preocupação compreensível de propor remédios concretos ao problema premente, em alguns dos aspectos éticos trabalhados, não mencionou a existência de posicionamentos divergentes.

Quanto à privacidade, por exemplo, o IBC, apesar de reconhecer que a proteção dos dados é necessária, afirma que os conceitos tradicionais de propriedade não são adequados aos dados, encarados no documento como um bem comum. A posição de uma parte importante do debate, que ainda busca mecanismos de recuperar e proteger tal propriedade dos dados como mecanismo de proteção da privacidade, não é mencionada.

Mesmo diante das promessas associadas ao fenômeno Big Data, a visão de que a privacidade é obsoleta não é hegemônica. A proteção e defesa da privacidade

continuam a ocupar um espaço central no debate bioético em torno de novas tecnologias, como é o caso das práticas data-cêntricas de saúde (89), mas também quanto a temas como a genômica e o uso de informações genéticas (90), e a neurociência (91), por exemplo. De nenhuma forma é consensual, na academia ou na percepção pública, a visão de que a privacidade é algo do qual devemos desistir.

O processo livre de decisão a respeito de como, com quem e o quanto se deseja compartilhar a esfera íntima, proporcionada pela existência da privacidade, nos permite ter relações próximas e significativas e também aquelas mais distantes. A existência da privacidade consente a gradação dos diferentes níveis de relação, das mais íntimas, às mais formais, segundo escolha que somente a preservação da privacidade permite. A privacidade, assim, é fundamental em todos os níveis: para a formação dos indivíduos, para o tecido social e, em última instância, para o pleno exercício democrático, uma vez que a vigilância prejudica a possibilidade de dissenso (92, 93, 94).

As violações da privacidade afetam, portanto, muito mais do que as vidas dos indivíduos. Sua defesa não é obsoleta, egoísta ou paradoxal.

Mesmo em momentos como a pandemia ainda tristemente vigente, o necessário aumento no fluxo de dados, com vistas a diminuir a propagação do vírus, tem um custo alto para a privacidade dos indivíduos e também para a coletividade com consequências futuras ainda imprevisíveis. A Organização Mundial de Saúde alertou que atentados contra direitos humanos e liberdades fundamentais são um risco. Usos discriminatórios e de cerceamento de liberdades e autonomia são algumas das preocupações. O órgão destacou que o mais sério risco, no entanto, é normalizar as medidas de violação da privacidade tomadas por motivação excepcional (95).

No relatório, também o exercício da autonomia se configura como um direito, ao menos em parte, permanentemente perdido. Isso porque a agregação de dados e a busca por relações e causalidades imprevistas, finalidades do uso de Big Data, impossibilitam uma manutenção total da privacidade e autonomia. O Relatório procura propor meios alternativos de consentimento, conforme já mencionado, como o *opt-out* ou o consentimento amplo. De forma geral, a manifestação do IBC reflete a compreensão de que tais perdas são irreversíveis, mas que elas gerarão ganhos em conhecimento e o compartilhamento dos benefícios em nível coletivo que justificam a diminuição desses direitos.

A discussão no Relatório, mas também em grande parte da literatura, parece girar em torno de uma contraposição entre os interesses individuais e coletivos. A conclusão costuma ser a de que o uso de Big Data se converte em benefício social. Nessa perspectiva, porém, residem dois problemas: primeiramente, é que, conforme já mencionado, a garantia de direitos como a autonomia e privacidade, mesmo que tratando-se de prerrogativas individuais, são estruturantes para a sociedade; finalmente, não está sempre demonstrado e inequívoco se os resultantes finais do uso de Big Data são de fato sempre benéficos para a sociedade. Observados os interesses comerciais gigantescos, a acumulação de riquezas observada nos últimos anos e a concentração de poder em figuras do mundo corporativo que controlam comportamentos sem o devido mandato democrático, esse benefício coletivo é um campo de disputa.

O Relatório é rico em recomendações concretas a respeito de uma regulamentação do uso de forma a evitar efeitos negativos do uso de Big Data. É sugerido à ONU a liderança no processo de governança internacional, a ser seguida por agências nacionais, do uso de dados; a Agência Internacional de Energia e as Agências de Meio Ambiente são chamadas a agir para economizar recursos ambientais raros; a OMS é chamada a controlar as aplicações em saúde e se recomenda a criação de um sistema de vigilância global para controlar o uso de dados em saúde; a Unesco é solicitada a promover uma convenção sobre a proteção sobre a privacidade e a OCDE recebe a recomendação de desenvolver uma estrutura de compartilhamento dos benefícios das aplicações de Big Data. Contudo, não se menciona uma recomendação que enfrente a modulação dos comportamentos dos usuários, que mitigue o perigo democrático no uso dessas tecnologias e que vigie as práticas de oligopólio e monopólio na comoditização de privacidade.

Quanto à modulação do comportamento, na literatura, é identificada a aguda capacidade de determinar comportamentos e percepções que se pode adquirir a partir do uso de dados. Essas mudanças que atingem as formas humanas de consumir, se relacionar, se expressar, enfim de estar no mundo, também a maneira de se compreender nele. Isso, conforme compreendem alguns autores (49), precisa ser estudado também em sua dimensão ética. Outros autores, como Arga (65), são mais enfáticos ao alertar que a capacidade preditiva tem sido usada com fins de controle e vigilância e levam a uma grande concentração de poder em nível global.

Embora haja um tom otimista, na literatura e no próprio Relatório do IBC, quanto aos efeitos coletivos do fenômeno Big Data, diversos autores expressam a preocupação com um aumento das desigualdades econômicas, com a exclusão de pessoas de baixa renda de oportunidades, e com a perpetuação de preconceitos e vieses presentes nos algoritmos (96). Ao falar de dados sensíveis, como informações sobre saúde, as discriminações resultantes da liberação de dados para companhias de seguro ou empregadores são bastante mencionadas (55, 97). Existe ainda o risco de que informações oriundas da prática de categorizar os indivíduos em perfis possam criar estigmas e prejudicar grupos de pacientes e cidadãos, devido a alguma doença, estilo de vida, ou qualquer fator que gere custos extras para as operadoras de saúde (76).

Adicionalmente, existe um risco aumentado de que a periferia do sistema, quer sejam partes ou nações inteiras, sirva apenas como “fazendas de dados” (98), ou seja, encarada apenas como fonte, local onde o dado é somente coletado como se fosse uma matéria-prima. Neste modelo que mercantiliza os dados, enquanto a *commodity* é extraída da periferia, a agregação de valor por meio da análise e os benefícios gerados pelo processo se destinam a regiões centrais. Os efeitos de perpetuação de injustiças são evidentes.

A busca pela justiça social é característica distintiva das análises na bioética latino-americana, principalmente na Bioética de Intervenção (17). É esse o movimento que impulsionou achar o que se considera o principal resultante dessa pesquisa, ou seja, a identificação de uma categoria/conceito que serve de chave interpretativa e permite a compreensão do fenômeno de Big Data e suas controvérsias éticas a partir da preocupação com a justiça. Trata-se de identificar os propósitos e mecanismos segundo os quais o uso de dados aglomerados opera a perpetuação da injustiça por meio de um processo de expropriação/acumulação de um poder/valor: isso se dá por meio da comoditização da privacidade.

6.3 A COMODITIZAÇÃO DA PRIVACIDADE

A comoditização da privacidade consiste no mecanismo central da era de vigilância do capitalismo. Este período, nas palavras de Zuboff (70), consiste em:

Uma nova ordem econômica que reivindica a experiência humana como matéria-prima gratuita para práticas comerciais dissimuladas de extração, previsão e vendas; Uma lógica econômica parasítica na qual a produção de bens e serviços é subordinada a uma nova arquitetura global de modificação de comportamento; Uma funesta mutação do capitalismo marcada por concentrações de riqueza, conhecimento e poder sem precedentes na história da humanidade; A estrutura que serve de base para a economia de vigilância; Uma ameaça tão significativa para a natureza humana no século XXI quanto foi o capitalismo industrial para o mundo natural nos séculos XIX e XX; A origem de um novo poder instrumentário que reivindica domínio sobre a sociedade e apresenta desafios surpreendentes para a democracia de mercado; Um movimento que visa impor uma nova ordem coletiva baseada em certeza total; Uma expropriação de direitos humanos críticos que pode ser mais bem compreendida como um golpe vindo de cima: uma destituição da soberania dos indivíduos (70, p. 7).

Um bom exemplo da mudança de paradigma na coleta e uso de dados pessoais para fins comerciais é o de casas automatizadas. Zuboff (70) exemplifica o assunto falando que nos anos 2000 um grupo de pesquisa colaborou para criar uma casa consciente, com sensores e computadores para tal. Eles entendiam que estavam iniciando um campo de conhecimento totalmente novo e, por conta disso, assumiram que os direitos e poder sobre esse conhecimento pertenceriam de forma exclusiva aos moradores. Cerca de duas décadas depois, casas inteligentes já são uma realidade, entretanto, cada peça de hardware que compõe o sistema de uma casa inteligente vem com uma política de privacidade, um contrato de termos de serviço e um contrato de licença para usuário final. Neles, os sujeitos formalmente abdicam das impressões digitais que suas vidas cotidianas geram. Como reforça Zuboff:

Esses documentos revelam consequências opressivas para a privacidade e a segurança, nas quais informações sensíveis do indivíduo e da casa são compartilhadas com outros dispositivos inteligentes, departamentos não identificados de corporações e terceiros, para propósitos de análise preditiva e vendas a outras partes não especificadas (70, p. 17).

Com isso, fica claro que os direitos têm sido usurpados e reivindicados de forma unilateral, às custas de experiências alheias e do conhecimento que flui destas (70).

Respondendo a uma lógica de mercado sem precedentes, a comoditização da privacidade se trata da utilização de mecanismos de vigilância para apropriação da privacidade, que passa a ser tratada como matéria-prima, commodity plena e indistintamente transacionável. Os dados são captados nas mais diferentes atividades, agregados e comercializados como aglomerados. Como uma commodity, a privacidade, apesar de ser pessoal, é vendida independentemente do indivíduo que

produz os dados. O que interessa é a capacidade de prever e induzir o comportamento, principalmente de consumo, de determinado segmento do mercado (70).

A gravidade do evento e suas possíveis implicações para os Direitos Humanos foi já alertada anteriormente e a singularidade do fenômeno é resumida de forma cristalina por Albuquerque:

“a captura da experiência humana que passa a ativo mercadológico é sem precedentes, igualmente, essa espécie de poder que emerge dos meios digitais do capitalismo de vigilância também é único na história” (9, p. 41).

Essa nova realidade de mercado afeta de forma ainda mais intensificada países periféricos, nos quais a população, além dos desafios apontados até aqui quanto ao Big Data, acumula injustiças diárias, situações de vulnerabilidade. Essas pessoas têm pouco ou nenhum mecanismo de, primeiramente, compreender o que lhe está sendo retirado e de, finalmente, resistir a isso.

Segundo Zuboff (70), “essa lógica transforma a vida comum na renovação diária de um pacto faustiano do século XXI”. Isso porque é quase impossível se livrar da forma que esses sistemas funcionam. Ainda mais quando ficamos dependentes de categorias como “monopólio” e “privacidade” no debate do tema, sendo que são termos insuficientes para identificar e debater os pontos mais cruciais e sem precedentes de tal regime.

O acesso democrático aos benefícios oriundos do desenvolvimento tecnológico, já se mostrava algo distante (99). Com a inserção do capitalismo de vigilância e sua nova lógica de mercado, o acesso democrático a todos os benefícios de Big Data se torna virtualmente impossível.

Muitos dados são coletados com a justificativa de que contribuirão para o aprimoramento de produtos e serviços. No entanto, o que se obtém é um superávit comportamental do proprietário. Esses dados são agregados e comercializados com a finalidade de alimentar avançados processos de “educação” de inteligência de máquina. Com o desenvolvimento do mercado da predição de comportamentos, as corporações, que se utilizavam de sobras dos dados para adivinhar preferências de consumos, passaram a usar complexos algoritmos para moldar os comportamentos de indivíduos em escala. Nada mais previsível do que comportamentos induzidos (70).

Na existência do capitalismo de vigilância, os dados não são apenas desviados de sua função original. São criadas vias de dados para que a experiência digital se

converta em combustível de análises e predições que, por sua vez, são utilizadas para fins lucrativos de quem controla esse sistema. Os usuários que produzem os valiosos dados, recebem em troca serviços “gratuitos” ou a promessa de uma experiência personalizada de consumo.

A manipulação e vigilância dos comportamentos têm se mostrado um negócio rentável. Quanto menos salvaguardas jurídicas existem, como no caso dos países periféricos, maior é a chance de acumulação de dados e, conseqüentemente, de enriquecimento de poucos às custas de muitos.

Coletar dados e utilizar estes para criar perfis específicos para anúncios comerciais não é mais o suficiente no capitalismo de vigilância, o próximo passo já vem sendo dado, trata-se de influenciar o comportamento de indivíduos de modo a maximizar o ganho econômico, conforme exposto no trecho a seguir:

Pressões de natureza competitiva provocaram a mudança, na qual processos de máquina automatizados não só conhecem nosso comportamento, como também moldam nosso comportamento em escala (70, p. 19).

Apesar da já substancial e expansiva intrusão na privacidade, a vigilância dos indivíduos é cada vez mais normalizada e aceita e tem se desdobrado em modulação comportamental massiva. A contínua vigilância, necessária para moldar comportamentos, representa um grande risco para a liberdade de expressão, de movimento e reunião. A consequência é uma profunda ameaça que paira ao mesmo tempo sobre as liberdades individuais e os sistemas democráticos.

A percepção pública destes riscos tem aumentado após recentes denúncias sobre o uso deletério de mídias digitais. Por meio de algoritmos, são criados filtros que, por meio de vigilância e modulação comportamental, criam bolhas que reforçam as cosmovisões de indivíduos e colaboram com o desenvolvimento de práticas racistas e extremistas, por exemplo (94).

Com o caso da *Cambridge Analytica*, compreendemos que esses efeitos são pretendidos para agir em escala global. A empresa, contratada para interferir no processo eleitoral dos Estados Unidos e no Reino Unido, lançava mão de mecanismos digitais de arquitetura de escolha, um termo mais ameno para a modulação comportamental (100).

É facilmente imaginável que em outro contexto de poder global, sem a disseminação de falsas notícias e negacionismo pelas mídias digitais, a própria

pandemia poderia ter tido um curso diferente. Essa é uma outra maneira com que a realidade de Big Data e circulação de dados começa a afetar a saúde das populações.

Ainda mais recentemente, com os chamados *facebook files*, tomamos conhecimento de programações destinadas a direcionar conteúdo apologético à anorexia e que tinham adolescentes do gênero feminino como alvo. Com o vazamento de documentos de grandes corporações do mundo digital, foi revelado que as corporações possuíam pleno conhecimento das negativas consequências para a saúde mental dos adolescentes expostos a redes sociais (101). Esse é um tema de interesse para a bioética, porque essa é uma nova maneira com que grupos inteiros da população, expostos a manipulação inconsciente de seu comportamento, passam a experimentar agravos que em breve constituirão importantes problemas de saúde pública.

A coleta de dados e informações pessoais vem sendo um debate de órgãos reguladores e é algo que divide opiniões. Apesar da vigilância tradicional, por assim dizer, aquela que envolve a gravação de áudio e vídeo por alguém ser algo repudiado e constantemente rechaçado judicialmente, caso feito sem autorização, a vigilância por parte de grandes corporações vem sendo realizada e aceita. Diversas corporações coletam dados muito além dos necessários para o funcionamento de seus aplicativos e, como já apresentado, utilizam-se disso para a criação de vias de dados nas quais é garimpado enorme lucro e poder.

É necessário que a dimensão digital detectada na análise da literatura realizada nesta pesquisa seja de fato percebida em sua extensão. A abordagem regulatória que foca meramente na proteção dos indivíduos, como faz inclusive o Relatório do IBC, não é suficiente para pautar todos esses desafios.

No contexto nacional, a regulação sobre proteção de dados, representada pela LGPD (102), apesar de ter esse nome, resguarda o segredo industrial e comercial acima do direito ao acesso aos dados pessoais em seu artigo 9º. O normativo prevê tantas isenções para a obrigação de obtenção de consentimento do proprietário dos dados, inclusive para realização de estudos e pesquisas, que nem mesmo essa preponderância dos interesses individuais pode ser afirmada. O estudo dessa legislação não é objeto desta dissertação, mas é premente que haja maiores estudos e debates bioéticos a respeito das plenas consequências da abordagem regulatória brasileira sobre dados pessoais.

Considerados todos esses aspectos, a dimensão estrutural que a comodificação inspira, caracterizando possível uma nova fase do capitalismo, não pode ser plenamente compreendida por uma oposição entre interesses individuais ou coletivos. Indivíduos e sociedade têm igual interesse em limitar a capacidade de modulação do comportamento por parte de grandes corporações. Esse é um risco para liberdades individuais, mas também para as democracias em função da cumulação de poder e riqueza. É importante realizar, no entanto, que esse risco atinge diversamente as populações. Enquanto centralmente o foco está na medicina personalizada e em possibilidades técnicas de anonimização e proteção de dados, na periferia, a vida cotidiana das pessoas vulneráveis se converte em fazendas de dados.

7 CONSIDERAÇÕES FINAIS

A presente dissertação buscou apresentar o atual estado da discussão sobre os desafios enfrentados na utilização de Big Data em saúde. A discussão atual revela que muitos dos desafios ainda não possuem o debate em níveis ideais, de modo que a literatura (mesmo especializada, como o Relatório do IBC sobre Big Data e Saúde (4) apenas menciona determinados desafios, deixando de evidenciar as controvérsias que o cercam.

Observou-se que muito da discussão e dos aspectos éticos identificados na literatura não são exclusivos do campo da saúde. As repercussões mais concretas dessas novas tecnologias sobre a saúde dos indivíduos, com repercussões sobre aspectos físicos e mentais, bem como suas interações com fatores ambientais, não foi tema abordado pela literatura estudada, ainda que provavelmente venham a se constituir temas de investigação ética no futuro.

A nula ou baixa produção científica nacional e regional (nacional e latino-americana) sobre as implicações éticas do uso de Big Data em saúde é ainda mais crítica quando se reflete sobre a injusta distribuição dos resultantes dessas tecnologias. Diante de todos os aspectos éticos aqui abordados, a ausência de discussão sobre Big Data e suas implicações, especialmente para a saúde, pode implicar na transformação regional em uma fazenda de dados, de modo que nenhum benefício real sobre o uso dos dados seja colhido e que apenas haja o fornecimento de insumos (dados) para análises localizadas em outros países, por sua vez melhores preparados para utilizar a tecnologia.

A dimensão de desafios mais relatada foi a social, o uso de dados em mídias sociais operou um rearranjo em muitas funções das vidas cotidianas e isso é percebido socialmente. A discussão sobre a injustiça e discriminação foi importante na literatura e no Relatório do IBC. A necessidade de compartilhamento de benefícios com países periféricos também foi tratada, mas a compreensão dos mecanismos que perpetuam a injustiça é limitada. Esta permanece enquanto não se evidenciam os interesses econômicos envolvidos na comodificação da privacidade. O conceito, proposto por Zuboff (70), serve de chave de interpretação crítica para o fenômeno, mas ainda permanece pouco explorado.

A dimensão econômica na literatura tratou mais de promover uma boa percepção pública, para promover a adesão à tecnologia, do que a conscientização plena da população. Em termos gerais, ainda que se apontem os aspectos éticos relacionados a dimensões econômicas e sociais, a visão prevalente é de que a tecnologia se traduz em bem comum, e a justiça deriva da garantia de compartilhamento dos benefícios. O fato que ela é desenvolvida por grandes corporações, que expropriam privacidade, acumulam poder e tem como propósito a geração de lucros crescentes, não é problematizado.

O comparativo dos desafios encontrados na busca na literatura com os aspectos éticos levantados no Relatório do IBC sobre Big Data e Saúde (4) se mostrou frutífero, dado que apesar de considerado um documento norteador, foi possível observar aspectos, como a condição humana, a comoditização da privacidade e a modulação comportamental, que o relatório não discutiu. Ainda, outros aspectos, menos citados pela literatura, são mencionados no relatório, como a preocupação ambiental com a infraestrutura necessária para que Big Data funcione, assim exemplificando a proposta ampla que o documento se propõe a trazer, apesar de não buscar ser exaustivo na temática.

A análise acerca da moralidade do uso de Big Data em Saúde, a partir da leitura do relatório e da literatura, revelou os focos do debate atual.

Dentre eles, destacou-se nesse estudo a leitura desse novo fenômeno tecnológico a partir da chave interpretativa da comodificação da privacidade. Por meio do referencial de Zuboff (70), nota-se que as novas tecnologias de tratamento de dados capturam a privacidade e transformam-na em ativo comercializável. Realiza-se que as grandes corporações que lideram essa nova fase têm como *métier* determinar comportamento de massas, sem qualquer tipo de mandato democrático. Não há porque esperar que em meio da busca de crescentes lucros, as tecnologias possam simples e automaticamente se traduzir em benefícios justos e amplamente distribuídos para a humanidade.

O problema no uso de Big Data não é o uso de dados pessoais. A partir deles organizam-se as políticas epidemiológicas e as práticas de saúde coletiva há pelo menos três séculos. Historicamente, a saúde pública, que se vale dos dados pessoais, constitui meio de mitigação de injustiça e promoção de qualidade de vida. A maneira como se converte esse tipo de agregação de dados pessoais em benefício coletivo, no entanto, não é um automatismo, mas uma consciência profunda da dimensão ética

dessa prática e um comprometimento com a promoção de bem-estar para toda a população e, principalmente, para os vulneráveis. Os benefícios coletivos que se esperam do uso de Big Data dependem de leituras críticas quanto à exploração dos indivíduos como meros produtores de dados e consumidores de produtos. O uso de agregados de dados demanda instâncias de controle social e de governança democrática que promovam uma luta constante para promoção da justiça social, compartilhamento dos benefícios e proteção dos vulneráveis. Esses resultados não vêm automaticamente do uso de dados, mas devem ser objeto de luta.

É premente que os aspectos éticos do uso de Big Data em saúde sejam refletidos em contextos latino-americanos e periféricos. Na Bioética de Intervenção, essa preocupação com o acesso aos benefícios das práticas de saúde se traduz na busca pela equidade. Essa comumente é associada às análises das situações persistentes, derivadas de desigualdades de acesso à saúde que não deveriam mais ocorrer (17). As injustiças sociais, no entanto, continuam atravessando os corpos vulneráveis também quando se apresentam as situações emergentes, impostas pelas novas tecnologias. No caso de Big Data, se a discussão não acontece a partir de uma perspectiva autóctone, preocupada com a justiça, o risco é que, em nome de um suposto bem comum, os corpos adoecidos periféricos continuem a gerar conhecimento, só para que esse se traduza em benefício em saúde para outros.

REFERÊNCIAS

- (1) Mittelstadt BD, Floridi L. The ethics of Big Data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics* 2016; **22**:303-341.
- (2) Al-Mekhlal M, Ali Khwaja A. A synthesis of Big Data definition and characteristics. In: 2019 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC). New York: IEEE, 2019. p. 314-322.
- (3) Demchenko Y, Grosso P, De Laat C, Membrey P. Addressing Big Data issues in scientific data infrastructure. 2013 International conference on collaboration technologies and systems 2013, pp. 48-55.
- (4) Unesco/International Bioethics Committee (IBC). Report of the IBC on Big Data and health. Paris: United Nations Educational, Cultural and Scientific Organization, 2017. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000248724> (Acessado em 15 de dezembro de 2021).
- (5) Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of Big Data in public health. *PLoS Computational Biology* 2015; 11(2):1-7.
- (6) Myers J, Frieden TR, Bherwani KM, Henning KJ. Ethics in public health research: privacy and public health at risk: public health confidentiality in the digital age. *American Journal of Public Health* 2008; 98(5):793-801.
- (7) Khoury MJ, Ioannidis JPA. Big data meets public health. *Science* 2014; 346(6213):1054-1055.
- (8) Davenport TH, Barth P, Bean R. How “Big Data” is different. *MIT Sloan Management Review* 2012; 54(1):22-24.
- (9) Albuquerque A. A privacidade na era dos algoritmos sob a ótica dos direitos humanos. In: Neto ABS, Almeida GHCB (org.). *Norma, linguagem e teoria do direito: reflexões para a compreensão do direito do século XXI*. Goiânia: Editora Espaço Acadêmico, 2020. p. 36-57.
- (10) ONU. Declaração Universal dos Direitos Humanos. 1948. Disponível em: <http://www.un.org/en/universal-declaration-human-rights/> (Acessado em 15 de dezembro de 2021)
- (11) United Nations Educational, Scientific and Cultural Organization (Unesco).

Declaração Universal sobre Bioética e Direitos Humanos. Tradução brasileira por Tapajós A, Prado MM, Garrafa V. Brasília: Cátedra Unesco de Bioética; 2005a.

- (12) Dussel E, Mendieta E, Bohórquez C. El pensamiento filosófico latinoamericano, del Caribe y "latino"(1300-2000): historia, corrientes, temas y filósofos. Siglo XXI; 2009.
- (13) Boyd D, Crawford K. Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 2012; 15(5):662-679.
- (14) Zwitter A. Big Data ethics. *Big Data & Society* 2014; 1(2):1-6.
- (15) Zuboff S. Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum* 2019a; 28(1):10-29
- (16) McCarthy MT. The Big Data divide and its consequences. *Sociology Compass* 2016; 10(12):1131-1140.
- (17) Garrafa V, Porto D. Intervention bioethics: a proposal for peripheral countries in a context of power and injustice. *Bioethics* 2003; 17(5-6):399-416.
- (18) Dias G, Vieira A. Big data: questões éticas e legais emergentes. *Ciência Da Informação* 2013; 42(2):174-184.
- (19) Bagheri A. Global bioethics: the impact of the Unesco international bioethics committee. New York: Springer, 2015.
- (20) Chen M, Mao S, Liu Y. Big data: a survey. *Mobile Networks and Applications* 2014; 19:171-209.
- (21) Russom P. Big data analytics. TDWI Best Practices Report, Fourth Quarter 2011; 19(4):1-34.
- (22) Davenport TH, Dyché J. Big data in big companies. *International Institute for Analytics* 2013; 3(1):1-31.
- (23) Chen CLP, Zhang CY. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Information Sciences* 2014; 275(1):314-347.
- (24) Yan J, Liu N, Yan S, Yang Q, Fan W, Wei W, et al. Trace-oriented feature analysis for large-scale text data dimension reduction. *IEEE Transactions on Knowledge and Data Engineering* 2010; 23(7):1103-1117.
- (25) Bezdek JC. Pattern recognition with fuzzy objective function algorithms. Boston: Springer, 1981.

- (26) Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials* 2016; 18(2):1153-1176.
- (27) Xindong W, Xingquan Z, Gong-Qing W, Wei D. Data mining with Big Data. *IEEE Transactions on Knowledge and Data Engineering* 2014; 26(1):97-107.
- (28) Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ. Big data for health. *IEEE journal of biomedical and health informatics* 2015; 19(4):1193-1208.
- (29) Garrafa V. Bioética. In: Giovanella L, Escorel S, Lobato LVC, Noronha JC, Carvalho AI (orgs.). *Política e Sistema de Saúde no Brasil*. Rio de Janeiro: Fiocruz/Cebes, 1a ed., 2008. p. 835-869.
- (30) Potter VR. Bioethics, the science of survival. *Perspectives in Biology and Medicine* 1970; 14(1):127-153.
- (31) Tribunals NM. Nuremberg Code. *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2*, pp. 181-182. Washington: U.S. Government Printing Office, 1949. Disponível em: <https://core.ac.uk/download/pdf/24066735.pdf> (Acessado em 16 de junho de 2021).
- (32) World Medical Association (WMA). Declaration of Helsinki. Recommendations guiding doctors in clinical research, 1964. Disponível em: <https://www.wma.net/wp-content/uploads/2018/07/DoH-Jun1964.pdf> (Acessado em 16 de junho de 2021).
- (33) Gamble VN. Under the shadow of Tuskegee: african americans and health care. *American Journal of Public Health* 1997; 87(11):1773-1778.
- (34) Beauchamp T. The principles approach. *The Hastings Center Report* 1993; 23(6): S9.
- (35) Junges J. Bioética como casuística e como hermenêutica. *Revista Brasileira de Bioética* 2005; 1(1):28-44.
- (36) Beauchamp TL, Childress JF. *Principles of biomedical ethics*. New York: Oxford University Press, 1979.
- (37) Clouser KD, Gert B. A critique of principlism. *Journal of Medicine and Philosophy* 1990; 15(2):219-236.
- (38) Azambuja LEO, Garrafa V. A teoria da moralidade comum na obra de Beauchamp e Childress. *Revista Bioética* 2015; 23(3):634-644.

- (39) Porto D, Garrafa V. Bioética de intervenção: considerações sobre a economia de mercado. *Revista Bioética* 2005; 13(1):111-123.
- (40) Garrafa V. Bioética das Situações Persistentes e das Situações Emergentes. *Fragmentos de Cultura* 1996; 6(21): 51-52.
- (41) Cunha T, Garrafa V. Vulnerability: a key principle for global bioethics? *Cambridge Quarterly of Healthcare Ethics* 2016; 25(2):197-208.
- (42) Packer AL, Tardelli AO, Castro RCF. A distribuição do conhecimento científico público em informação, comunicação e informática em saúde indexado nas bases de dados MEDLINE e LILACS. *Ciência & Saúde Coletiva* 2007; 12(3):587-599.
- (43) BIREME / OPAS / OMS (Brasil) Acesso às fontes de informação da Biblioteca Virtual em Saúde. São Paulo: BIREME / OPAS / OMS, 2008. 23 p.
- (44) Souza MAR, Wall ML, Thuler ACMC, Lowen IMV, Peres AM. O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas. *Revista da Escola de Enfermagem da USP* 2018; 52(1):1-7.
- (45) Salviati ME. Manual do aplicativo IRaMuTeQ: compilação, organização e notas. 2017. Disponível em: <http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>. (Acessado em 15 de Dezembro de 2020).
- (46) Camargo BV, Justo AM. IRaMuTeQ: um software gratuito para análise de dados textuais. *Temas em Psicologia* 2013; 21(2):513-518.
- (47) Nascimento A, Menandro P. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. *Estudos e Pesquisas em Psicologia* 2006; 6(2):72-88.
- (48) Kronberger N, Wagner W. Palavras-chave em contexto: análise estatística de textos. In: Bauer MW, Gaskell G. *Pesquisa qualitativa com texto, imagem e som*. Petrópolis: Editora Vozes, 2002. p. 416-441.
- (49) Überall M, Werner-Felmayer G. Integrative biology and big-data-centrism: mapping out a bioscience ethics perspective with a S.W.O.T. Matrix. *OMICS: A Journal of Integrative Biology* 2019; 23(8):371-379.
- (50) Peirce AG, Elie S, George A, Gold M, O'Hara K, Rose-Facey W. Knowledge development, technology and questions of nursing ethics. *Nursing Ethics*

2020; 27(1):77-87.

- (51) Cesare N, Lee H; McCormick T, Spiro E, Zagheni E. Promises and pitfalls of using digital traces for demographic research. *Demography* 2018; 55(5):1979-1999.
- (52) Beier K, Schweda M, Schicktanz S. Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research. *BMC Medical Informatics and Decision Making* 2019; **19**:90.
- (53) Lawler M, Maughan T. From Rosalind Franklin to Barack Obama: data sharing challenges and solutions in genomics and personalised medicine. *The New Bioethics* 2017; 23(1):64-73.
- (54) Hallowell N, Parker M, Nellåker C. Big data phenotyping in rare diseases: some ethical issues. *Genetics in Medicine* 2019; 21(2):272-274.
- (55) Manrique de Lara A, Peláez-Ballestas I. Big data and data processing in rheumatology: bioethical perspectives. *Clinical Rheumatology* 2020; 39(4):1007-1014.
- (56) Canales C, Lee C, Cannesson M. Science without conscience is but the ruin of the soul: the ethics of Big Data and artificial intelligence in perioperative medicine. *Anesthesia and Analgesia* 2020; 130(5):1234-1243.
- (57) Middleton A. Society and personal genome data. *Human Molecular Genetics* 2018; 27(1):8-13.
- (58) Tossas-Milligan KY, Winn RA. Breaking the cycle of health inequities: the bioethics of data. *Journal of Health Care for the Poor and Underserved* 2019; 30(5):86-90.
- (59) Bassan S, Harel O. The ethics in synthetics: statistics in the service of ethics and law in health related research in Big Data from multiple sources. *Cleveland-Marshall Journal of Law and Health* 2018; 31(1):87-117.
- (60) Evans BJ. Power to the people: data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law* 2017; 19(2):243-265.
- (61) Ienca M, Ignatiadis K. Artificial intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB Neuroscience* 2020; 11(2):77-87.
- (62) Wang S, Bonomi L, Dai W, Chen F, Cheung C, Bloss CS, et al. Big Data Privacy in Biomedical Research. *IEEE Transactions on Big Data* 2020; 6(2):296-

308.

- (63) Laviolle B, Denèfle P, Gueyffier F, Bégué É, Bilbault P, Espérou H, et al. The contribution of genomics in the medicine of tomorrow, clinical applications and issues. *Therapies* 2019; 74(1):9-15.
- (64) Dhar V. Should we regulate digital platforms? *Big Data* 2017; 5(4):277-278.
- (65) Arga KY. COVID-19 and the futures of machine learning. *OMICS: a journal of integrative biology* 2020; 24(9):512-514.
- (66) Goodman KW. Ethics in health informatics. *Yearbook of Medical Informatics* 2020; 29(1):26-31.
- (67) Allen J, Adams C, Flack F. The role of data custodians in establishing and maintaining social licence for health research. *Bioethics* 2019; 33(4):502-510.
- (68) Robinson K. A false promise of COVID-19 'big' health data? Health data integrity and the ethics and realities of Australia's health information management practice. *Health Information Management Journal* 2020; 50(1-2):9-12.
- (69) Rothstein MA. Ethical issues in Big Data health research: currents in contemporary bioethics. *Journal of Law, Medicine & Ethics* 2015; 43(2):425-429.
- (10) Zuboff S. A era do capitalismo de vigilância: a luta por um futuro humano na fronteira do poder. Rio de Janeiro: Intrínseca, 2019b.
- (71) Dove ES, Özdemir V. What role for law, human rights, and bioethics in an age of Big Data, consortia science, and consortia ethics? The importance of trustworthiness. *Laws* 2015; 4(3):515-540.
- (72) French R. What is the good of it — ethical controls of human subject health research? *Journal of Bioethical Inquiry* 2018; 15(4):589-602.
- (73) Lajaunie C, Ho CWL. Pathogens collections, biobanks and related-data in a one health legal and ethical perspective. *Parasitology* 2018; 145(5):688-696.
- (74) Spector-Bagdady K, Jagsi R. Big data, ethics, and regulations: Implications for consent in the learning health system. *Medical Physics* 2018; 45(10):e845–847.
- (75) Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of Big Data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Medicine* 2019; 17(1):143–147.
- (76) Dorey CM, Baumann H, Biller-Andorno N. Patient data and patient rights: swiss

- healthcare stakeholders' ethical awareness regarding large patient data sets – a qualitative study. *BMC Medical Ethics* 2018; 19(1):1-14.
- (77) Rosen DL, Buchbinder M, Juengst E, Rennie S. Public health research, practice, and ethics for justice-involved persons in the Big Data era. *American Journal of Public Health* 2020; 110(S1):S37-38.
- (78) Lawler M, Morris AD, Sullivan R, Birney E, Middleton A, Makaroff L, et al. A roadmap for restoring trust in Big Data. *The Lancet Oncology* 2018; 19(8):1014-1015.
- (79) Margetts H, Dorobantu C. Rethink government with AI. *Nature* 2019; 568:163–5.
- (80) Cassel C, Bindman A. Risk, benefit, and fairness in a Big Data world. *The Journal of the American Medical Association* 2019; 322(2):105-106.
- (81) Sohrabi C, Mathew G, Franchi T, Kerwan A, Griffin M, Del Mundo JSC, et al. Impact of the coronavirus (COVID-19) pandemic on scientific research and implications for clinical academic training – a review. *International Journal of Surgery* 2021; 86(1):57-63.
- (82) Scimago Journal and Country Rank (SJR). International Science Ranking: all subject categories. 1996-2020. Disponível em: <https://www.scimagojr.com/countryrank.php>. (Acessado em 01 de Dezembro de 2021).
- (83) Garrafa V. Da bioética de princípios a uma bioética interventiva. *Revista Bioética* 2005; 13(1):125-134.
- (84) Garrafa V. Ampliação e politização do conceito internacional de bioética. *Revista Bioética* 2012; 20(1):9-20.
- (85) Harayama RM. Reflexões sobre o uso do Big Data em modelos preditivos de vigilância epidemiológica no Brasil. *Cadernos Ibero-Americanos de Direito Sanitário* 2020; 9(3):153-165.
- (86) Demotes-Mainard J, Cornu C, Guérin A, Bertoye PH, Boidin R, Bureau S, et al. How the new European data protection regulation affects clinical research and recommendations? *Therapies* 2019; 74(1):31-42.
- (87) Martins TGS, Costa ALFA, Martins TGS. Big data use in medical research. *Einstein (São Paulo)* 2018; 16(3):1-2.
- (88) Fulgêncio C, Nascimento W. Bioética de intervenção e justiça: olhares desde o sul. *Revista Brasileira de Bioética* 2012; 8(1-4):47-56.
- (89) Mehta N, Pandit A. Concurrence of Big Data analytics and healthcare: A

- systematic review. *International journal of medical informatics* 2018; 114(1): 57-65.
- (90) Cech, M. Genetic privacy in the “Big Biology” era: the “autonomous” human subject. *Hastings Law Journal* 2019; 70(3):851-886.
- (91) Leefmann J, Levallois C; Hildt E. Neuroethics 1995–2012: A bibliometric analysis of the guiding themes of an emerging research field. *Frontiers in human neuroscience*, 2016; 10(1): 1-19.
- (92) Pedersen DM. Psychological functions of privacy. *Journal of Environmental Psychology* 1997; 17(2):147-156.
- (93) Roessler B, Mokrosinska D. Privacy and social interaction. *Philosophy & Social Criticism* 2013; 39(8):771-791.
- (94) Cohen JE. What privacy is for. *Harvard Law Review* 2012; 126:1904-1933.
- (95) OMS. Joint Statement on Data Protection and Privacy in the COVID-19 Response. 2020.
- (95) Madden M, Gilman M, Levy K, Marwick A. Privacy, poverty, and Big Data: a matrix of vulnerabilities for poor americans. *Washington University Law Review* 2017; 95(1):53-125.
- (97) Gürsoy G, Harmanci A, Tang H, Ayday E, Brenner SE. When Biology Gets Personal: Hidden Challenges of Privacy and Ethics in Biological Big Data. *Pacific Symposium on Biocomputing* 2019, p. 386-390.
- (98) Carr M, Llanos JT. Data: Global governance Challenges. In: Weiss, Thomas G., and Rorden Wilkinson (ed.). *Global Governance Futures*. London: Routledge, 2021. digital file (sp). <https://doi.org/10.4324/9781003139836>
- (99) Garrafa V, Costa SIF, Oselka G. A bioética no século XXI. *Revista Bioética* 2009; 7(2):1-5.
- (100) Bartlett J. *The people vs tech: how the internet is killing democracy (and how we save it)*. London: Random House. 2018.
- (101) Horwitz J, Wells G, Seetharaman D, et al. Facebook files series. *New York: Wall Street Journal*, 2021. Disponível em: <https://www.wsj.com/articles/the-facebook-files-11631713039?> (Acessado em 15 de dezembro de 2021).
- (102) BRASIL. Lei n. 13709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais. *Diário Oficial da União, Brasília* (2018 ago. 15). Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-

2018/2018/lei/l13709.htm

ANEXO A – PROPOSTA DE TRADUÇÃO DO AUTOR DO “RELATÓRIO DO IBC SOBRE BIG DATA E SAÚDE”

RELATÓRIO DO IBC SOBRE BIG DATA E SAÚDE

No âmbito de seu programa de trabalho para 2016-2017, o Comitê Internacional de Bioética da Unesco (IBC, da sigla em inglês) decidiu abordar o tema Big Data e saúde, incluindo, mas não se limitando às questões de autonomia, consentimento, proteção de dados, governança, etc.

Na 22ª Sessão (Ordinária) do IBC, em setembro de 2015, o Comitê estabeleceu um Grupo de Trabalho para desenvolver uma reflexão inicial sobre este tópico. O Grupo de Trabalho do IBC, utilizando o intercâmbio de e-mails, começou a preparar um texto sobre esta reflexão entre outubro de 2015 e maio de 2016. Ele também se reuniu em Colônia em maio de 2016 para refinar a estrutura e o conteúdo de seu texto. Com base neste trabalho, o Grupo de Trabalho do IBC preparou uma minuta preliminar de relatório que foi discutida durante sua 23ª sessão (Ordinária) em setembro de 2016. Como continuação desta discussão, o Grupo de Trabalho do IBC começou a revisar a minuta preliminar do relatório entre setembro e dezembro de 2016. O Grupo de Trabalho do IBC reuniu-se na Espanha em março de 2017 para aperfeiçoar ainda mais o texto. Um texto revisado na forma de uma minuta de relatório foi apresentado ao IGBC, ao IBC e ao COMEST entre maio e junho de 2017 para comentários. A minuta do relatório foi então revisada com base nos comentários recebidos. A minuta final do relatório foi discutida e revisada durante a 24ª sessão (Ordinária) do IBC, e foi adotada pelo Comitê em 15 de setembro de 2017.

Este documento não pretende ser exaustivo e não representa necessariamente o ponto de vista dos Estados-Membros da Unesco.

I. ESCOPO E DEFINIÇÕES

II. REGULAMENTAÇÃO LEGAL

III. VISÕES, TENDÊNCIAS E DESAFIOS

IV. ASPECTOS ÉTICOS

IV.1. Autonomia e Consentimento

IV.2. Privacidade e Confidencialidade

IV.3. Da propriedade à custódia e compartilhamento de benefícios

IV.4. Justiça

IV.4.1. Lacuna digital

IV.4.2. Não-discriminação

IV.5. Sustentabilidade em relação à energia e ao meio ambiente

IV.6. Pesquisa

V. GOVERNANÇA

VI. RECOMENDAÇÕES

BIBLIOGRAFIA

I. ESCOPO E DEFINIÇÕES

1. A existência digital de todos os tipos de dados está levando a um fenômeno em exponencial avanço chamado Big Data. Já estão sendo alcançadas e transformadas todas as dimensões da vida humana em todo o mundo. Neste relatório, o IBC se concentra em questões relevantes para a área da saúde a nível individual e público. Ele apresenta recomendações para que todo o potencial de Big Data possa ser explorado, garantindo ao mesmo tempo que a dignidade humana, os direitos fundamentais e as liberdades fundamentais sejam mantidos de acordo com o Artigo 3 da Declaração Universal sobre Bioética e Direitos Humanos (2005).

2. Big Data é caracterizado pelos chamados 5 Vs:

- a. *Volume* se refere à gigantesca quantidade de dados digitais. Ela está crescendo exponencialmente. Enquanto três quartos dos dados eram analógicos em 2000, hoje mais de 99% de todos os dados são dados digitais. Para 2020, a quantidade estimada de dados foi de 44 zetabytes (10^{21}) de dados digitais, para 2025 é de 180 zetabytes (IDC, 2014). Mas o volume de dados sozinho, vindo por exemplo de um sequenciamento genético completo, não se configura como Big Data.
- b. *Variedade* aponta para o fato de que existem diferentes tipos de dados provenientes de diversos tipos de fontes. Para fins de cuidados e pesquisa na área de saúde, várias fontes são relevantes: dados médicos individuais de pacientes, dados de saúde pública, dados de diferentes seguros de saúde, dados de pesquisa coletados por pesquisadores e cientistas cidadãos¹, por instituições privadas ou pelos próprios indivíduos, dados sobre estilo de vida, como por exemplo, aqueles vindos de aplicativos de saúde, dados de redes sociais e dados de comportamento de consumo. Estes dados podem ser classificados de diferentes maneiras e de acordo com diferentes critérios: existem, por exemplo, dados pessoais, dados anonimizados, metadados, dados primários e secundários.
- c. *Velocidade* refere-se à altíssima velocidade com a qual os dados podem ser coletados e processados. O rastreamento em tempo real e as soluções em nuvem permitem um processamento abrangente em segundos, produzindo até mesmo recomendações imediatas, por exemplo, para medicação, comportamento ou nutrição.
- d. *Validade* refere-se à qualidade dos dados e se eles realmente demonstram o que pretendem em relação ao conteúdo e à precisão. O contexto dos dados tem um papel importante aqui.
- e. *Valor* finalmente chama a atenção para o significado que os dados têm para uma determinada questão, por exemplo, em relação a uma determinada

¹ Neste relatório, o termo ciência cidadã refere-se a projetos de pesquisa que são conduzidos por leigos, frequentemente cooperando com ou sendo conduzidos por cientistas profissionais ou instituições de pesquisa.

doença. Aqui novamente é importante levar em conta o contexto dos dados.

3. Neste contexto, o IBC usa o termo Big Data na área da saúde denotando grandes coleções de dados complexos, provenientes de várias fontes, relacionados à saúde. Tipicamente, tais conjuntos de dados cobrem um número muito grande de indivíduos, mas a análise de todos os dados disponíveis de um único paciente sob certas condições também pode ser considerada uma análise de Big Data.

4. Big Data não pode ser processada com recursos tradicionais, mas requer maior poder computacional e o desenvolvimento de novos algoritmos. Diferentes tipos de algoritmos implicam em diferentes tipos de desafios éticos. Existem, por exemplo, algoritmos específicos para tarefas que são programados como uma sequência definida de operações claras, permitindo assim a transparência técnica. Mas existem também algoritmos na área de inteligência artificial, aprendizado de máquinas e aprendizagem profunda em que a transparência é difícil porque os algoritmos são ensinados com grandes conjuntos de dados e realizam operações mais ou menos "por si mesmos".

5. Não existe uma definição universal de saúde que poderia ser aplicada em todos os contextos de aplicação a todos os períodos da vida e em todas as áreas do mundo. As definições variam de acordo com a perspectiva (um entendimento subjetivo ou conceito objetivo) e de acordo com a finalidade da definição em termos de ações terapêuticas e/ou preventivas que podem então ser legitimadas. Por exemplo, a definição ampla da Organização Mundial da Saúde (OMS) é entendida como uma ideia reguladora guia para promover a saúde global. Ela define saúde como o estado de completo bem-estar físico, mental e social e não apenas a mera ausência de doença ou enfermidade, abrangendo assim toda a vida de um indivíduo em todos os aspectos. Quando os sistemas nacionais de saúde têm que especificar e limitar as responsabilidades dos profissionais e instituições médicas, é utilizada uma abordagem mais restrita.

6. A definição da OMS mostra que tudo na vida é importante no que se refere à saúde. Neste sentido, ela confirma uma abordagem holística da saúde e embarça a linha entre saúde no campo médico e estilo de vida. Big Data oferece a oportunidade técnica de apoiar tal visão holística, ao mesmo tempo em que suscita sérias preocupações sobre a proteção dos direitos humanos.

7. Em tempos de tecnologias da informação e comunicação (TICs), novos termos como eHealth (saúde eletrônica) e mHealth (saúde móvel) surgiram. A OMS define eHealth como "o uso de tecnologias de informação e comunicação para a saúde" (WHO, sem data), e mHealth como um subcomponente da eHealth em termos de "prática médica e de saúde pública apoiada por dispositivos móveis, tais como telefones celulares, dispositivos de monitoramento de pacientes, assistentes pessoais digitais (PDAs), e outros dispositivos sem fio" (WHO, 2011).

8. Há um número crescente de aplicativos no dia-a-dia (de acordo com estimativas, já existem mais de 300.000) que tratam de questões relacionadas à saúde. Independentemente da impossibilidade de controlar e garantir a qualidade de todas essas aplicações, dificilmente é possível delinear claramente as aplicações médicas de aplicações de saúde não médicas. Elas se enquadram principalmente em três categorias: primeiro, aplicações relacionadas ao tratamento, segundo, aplicações relacionadas à previsão e prevenção, e terceiro, aplicações relacionadas ao estilo de vida. Portanto, ao abordar Big Data e a saúde neste relatório, isto implica necessariamente mais do que os cuidados de saúde tradicionais e a pesquisa em saúde.

II. REGULAMENTAÇÃO LEGAL

9. Não há regulamentação específica do fenômeno de Big Data nos quadros jurídicos nacionais e internacionais. No entanto, existe um marco regulatório completo para a proteção de dados pessoais em muitas jurisdições legais, principalmente na Europa, do qual muitas regras podem ser aplicáveis na área de Big Data, embora seja uma nova realidade no sentido de quantidade, análise, acessibilidade e aplicação. Além disso, os países que não possuem leis específicas sobre proteção de dados pessoais ainda podem utilizar disposições constitucionais e legais, bem como princípios de direito comum, para o mesmo fim. Um bom exemplo a este respeito é a maioria dos países da Comunidade das Nações (UNCTAD, 2016). Portanto, não há falta de regulamentação, mas de disposições específicas e talvez de novos princípios que sejam adequados para regular as novas características de Big Data.

10. No quadro jurídico internacional, o Artigo 12 da Declaração Universal dos Direitos Humanos, adotada pela Assembleia Geral das Nações Unidas (ONU) em 1948, cobre a privacidade, afirmando que: "Ninguém será sujeito à interferência na sua vida privada, na sua família, no seu lar ou na sua correspondência, nem a ataque a sua honra e reputação. Todo ser humano tem direito à proteção da lei contra tais interferências ou ataques." (UN, 1948). Em um sentido semelhante, o artigo 8(1) da Convenção Europeia de Direitos Humanos estabelece que: "todos têm o direito ao respeito por sua vida privada e familiar, seu lar e sua correspondência" (CoE, 1950). O artigo 8(2) acrescenta que: "não deve haver ingerência de uma autoridade pública no exercício deste direito, exceto se estiver de acordo com a lei e for necessário numa sociedade democrática no interesse da segurança nacional, da segurança pública ou do bem-estar econômico do país, para a prevenção da desordem ou do crime, para a proteção da saúde ou da moral, ou para a proteção dos direitos e liberdades de outrem" (CoE, 1950).

11. Muitas das regulamentações a nível internacional foram desenvolvidas no contexto de fluxos internacionais de dados principalmente devido ao comércio em serviços de saúde, o que leva a transferências de dados transfronteiriças. As regulamentações de proteção de dados e fluxos internacionais de dados (2016) da

Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD, da sigla em inglês) são notáveis a este respeito. O acordo sobre os princípios centrais pode ser atribuído às Diretrizes da Assembleia Geral das Nações Unidas para a Regulamentação de Arquivos de Dados Pessoais Informatizados (1990), que contém princípios para garantir a proteção da privacidade e confidencialidade que, no mínimo, devem ser previstos nas legislações nacionais. Estes são os princípios de especificação de propósito e segurança. As diretrizes também exigem que os países designem uma autoridade que supervisione a observância desses princípios, sancione aqueles que violarem e prescreve a necessidade de proteger a privacidade durante o movimento transfronteiriço de dados pessoais. As diretrizes foram concebidas para reger arquivos informatizados e manuais que contenham informações pessoais (ver parágrafo 10 das diretrizes), mas os princípios ainda podem ser aplicados, em certa medida, no contexto de Big Data. Outras diretrizes não juridicamente vinculativas, que moldaram as estruturas jurídicas nacionais, são a Declaração de Taipei sobre Considerações Éticas em Relação aos Bancos de Dados de Saúde e Biobancos da Associação Médica Mundial (2016) e a Declaração de Helsinque (2013). Embora haja um amplo consenso sobre os princípios centrais de proteção de dados no coração da maioria das leis nacionais e regimes internacionais, o principal desafio é a divergência na implementação destes princípios, bem como nas leis detalhadas de proteção de dados do mundo (UNCTAD, 2016).

12. A Organização para Cooperação e Desenvolvimento Econômico (OECD, da sigla em inglês) desenvolveu suas Diretrizes para a Proteção da Privacidade e Fluxo Transfronteiriço de Dados Pessoais (2013). Há também um relatório da OECD que destaca a necessidade de desenvolvimento de estruturas legais apropriadas para o compartilhamento de informações (OECD, 2010). Este relatório destaca a necessidade de uma nova estrutura legal que permita o compartilhamento de informações relacionadas à saúde entre profissionais da saúde dentro e ao longo de organizações de saúde, bem como ao longo de fronteiras organizacionais e geográficas. O relatório observa que muito poucos países em sua área de competência abordaram realmente estes desafios. Mais recentemente, em janeiro de 2017, a OECD publicou sua Recomendação sobre Governança de Dados de Saúde. Eles destacam a importância de complementar a proteção legal de dados ao longo da educação e conscientização, do desenvolvimento de habilidades e da promoção de medidas técnicas. A Recomendação conclama os países a desenvolver e implementar estruturas de governança de dados de saúde que garantam a privacidade e, ao mesmo tempo, permitam o uso de dados de saúde que sejam de interesse público, de acordo com doze princípios de alto nível.

13. Com relação a todas as regulamentações, uma distinção importante deve ser feita entre a Europa e os EUA. A abordagem europeia é baseada na visão de que a privacidade é um direito humano fundamental e envolve uma regulamentação de cima para baixo e a imposição de regras gerais que restringem o uso de dados ou exigem consentimento explícito para esse uso. Os Estados Unidos, em contraste, empregam

uma abordagem setorial que se concentra na regulamentação de riscos específicos de danos à privacidade em contextos particulares, como cuidados de saúde e crédito. Isto coloca menos regras amplas sobre o uso de dados, permitindo que a indústria seja mais inovadora em seus produtos e serviços, ao mesmo tempo em que, às vezes, deixa também usos potenciais não regulamentados de informações que ficam entre diferentes setores. (USA, 2014)

14. Abordagens diferentes nas duas regiões levaram ao projeto da Estrutura de Proteção de Privacidade UE-EUA e Suíça-EUA (o escudo de privacidade) pelo Departamento de Comércio dos EUA, a Comissão Europeia e a Administração Suíça com o objetivo de garantir o cumprimento das exigências de proteção de dados pessoais no comércio transatlântico (USA, sem data). O escudo de privacidade substituiu o anterior Acordo Porto Seguro de 2000 e está sujeito a revisão anual para assegurar sua atualidade a medida que a tecnologia muda e o regime de proteção de dados da UE é transformado (cf. parágrafo 27 abaixo) (Weiss e Archick, 2016).

15. O Conselho da Europa aprovou em 1981 a Convenção para a Proteção de Indivíduos com Relação ao Processamento Automático de Dados Pessoais. O objetivo era proteger o indivíduo contra abusos que pudessem estar associados à coleta e ao processamento de dados pessoais. Posteriormente, o Conselho da Europa aprovou um Protocolo Adicional a esta Convenção (CoE, 2001) relativo às autoridades de supervisão e fluxos de dados transfronteiriços. Ele prevê a criação de autoridades supervisoras nacionais responsáveis por garantir o cumprimento das leis ou regulamentos adotados em conformidade com a Convenção. Além disso, os dados só podem ser transferidos se o Estado receptor ou organização internacional for capaz de fornecer um nível adequado de proteção. Em 2017, as Diretrizes do Conselho da Europa sobre a proteção de indivíduos com relação ao processamento de dados pessoais em um mundo de Big Data foram estabelecidas com base nesta Convenção.

16. A UE tem apenas uma competência legal limitada em matéria de saúde, que pode ser usada principalmente para promover a cooperação e a coordenação entre os Estados-Membros. Entretanto, na área de proteção de dados existe uma regulamentação comum ao longo da Diretiva 95/46/CE do Parlamento Europeu e do Conselho de 24 de outubro de 1995 relativa à proteção das pessoas físicas no que se refere ao processamento de dados pessoais e à livre circulação de tais dados (EU, 1995). Esta situação mudará em 2018. Em vez de uma Diretiva, haverá um Regulamento que é diretamente obrigatório em todos os Estados-Membros e é relevante também para as corporações extra europeias (EU, 2016).

- a. O novo Regulamento prevê que "o processamento de dados pessoais deve ser projetado para servir à humanidade. O direito à proteção de dados pessoais deve ser considerado em relação a sua função na sociedade e deve ser equilibrado com outros direitos fundamentais. Este Regulamento respeita todos os direitos fundamentais, liberdades e princípios reconhecidos na Carta como consagrados nos Tratados, em particular o

respeito à vida privada e familiar, ao lar e às comunicações, à proteção de dados pessoais, à liberdade de pensamento, consciência e religião, à liberdade de expressão e informação, à liberdade de conduzir um negócio, ao direito a um recurso efetivo e a um julgamento justo, bem como à diversidade cultural, religiosa e linguística" (EU, 2016, parágrafo 4).

- b. Em relação ao consentimento, o Regulamento estabelece que ele deve ser dado por um ato afirmativo claro. Se o consentimento da pessoa interessada for dado após um pedido por meios eletrônicos, o pedido deve ser claro, conciso e não desnecessariamente perturbador do uso do serviço para o qual é prestado. Mais tarde, acrescenta que "consentimento não deve ser considerado como dado livremente se o envolvido não tiver escolha genuína ou livre ou não puder recusar ou retirar o consentimento sem prejuízo" (EU, 2016, parágrafo 42).
- c. Na área de pesquisa, o Regulamento comenta que, "desde que sejam mantidos padrões éticos reconhecidos para a pesquisa científica, os sujeitos dos dados devem ser autorizados a dar seu consentimento a certas áreas de pesquisa científica", uma vez que "a finalidade específica do processamento de dados pessoais muitas vezes não pode ser totalmente identificada no momento da coleta de dados" (EU, 2016, parágrafo 33).
- d. Na área de saúde pública, o novo Regulamento menciona que "o tratamento de categorias especiais de dados pessoais pode ser necessário por razões de interesse público sem o consentimento do envolvido. Tal tratamento deve ser sujeito a medidas adequadas e específicas para proteger os direitos e liberdades das pessoas físicas" (EU, 2016, parágrafo 54).

17. Os Estados Unidos da América também oferecem um marco legal ao longo de algumas leis recentes. Não há regulamentação que diga respeito a Big Data, mas as corporações que realizam grandes operações de processamento de dados na área de saúde precisam seguir regulamentações de proteção de dados a nível federal: a Lei de Portabilidade e Responsabilidade de Seguros de Saúde (HIPAA, da sigla em inglês) e a Regra de Privacidade HIPAA (USA, 1996; USA, 2002). Ambos os regulamentos exigem salvaguardas apropriadas para proteger a privacidade das informações pessoais de saúde e estabelecem limites e condições sobre os usos e divulgações que podem ser feitos de tais informações sem a autorização do paciente. A regra também confere aos pacientes direitos sobre suas informações de saúde, incluindo direitos de examinar e obter uma cópia de seus registros de saúde, e de solicitar correções. O Ato de Recuperação e Reinvestimento Americano (ARRA, da sigla em inglês) e o Ato de Tecnologia de Informação da Saúde para Economia e Saúde Clínica (HITECH, da sigla em inglês) criam incentivos significativos para um maior uso de registros de saúde eletrônicos e também contém algumas regras sobre proteção de dados, principalmente sobre medidas relativas a violações de informações de saúde (USA, 2009).

18. A União Africana (AU, da sigla em inglês) e a Cooperação Econômica Ásia-Pacífico (APEC, da sigla em inglês) são outros exemplos de organizações regionais que desenvolveram estruturas legais regionais para a proteção de dados pessoais e privacidade. A Convenção da AU sobre Segurança Cibernética e Proteção de Dados Pessoais exige que o processamento de dados pessoais envolvendo informações genéticas e pesquisas em saúde seja realizado com a autorização da autoridade nacional de proteção (AU, 2014, Artigo 10[4][a]). A Estrutura de Privacidade atualizada da APEC prevê "um mecanismo multilateral que permite às Autoridades de Aplicação da Privacidade na região da APEC cooperar na aplicação transfronteiriça das Leis de Privacidade" (APEC, 2015).

19. Com relação à proteção de dados, autonomia e privacidade, o conceito legal de propriedade também é importante. Deve ser feita uma distinção clara entre o que constitui uma descoberta e o que constitui uma invenção. Normalmente, as novas invenções são patenteáveis, pois implicam uma atividade inventiva e são utilizadas industrialmente. Em 3 de julho de 2012, o Tribunal de Justiça Europeu (TJE) publicou sua decisão referência em *Used Soft GmbH contra Oracle International Corp.* (C-128/11). A decisão implica que há uma propriedade específica atribuída a bens intangíveis como software baixado via Internet. Embora a aplicabilidade deste modelo a outros bens digitais ainda deva ser considerada em futuras decisões judiciais, o TJE abriu a porta para uma discussão sobre a propriedade de bens intangíveis (Hoeren, 2014). Além disso, em um recurso do julgamento de *Football Dataco Ltd contra Brittens Pools Ltd*, de abril de 2010, o Tribunal de Justiça Europeu (TJE) enfatizou que o objetivo da Diretriz de Banco de Dados é o de: "estimular a criação de sistemas de armazenamento e processamento de dados a fim de contribuir para o desenvolvimento de um mercado de informação [...] e não para proteger a criação de materiais capazes de serem coletados em um banco de dados" (parágrafo 34).

20. No nível internacional, o acordo TRIPS, artigo 10(2), protege compilações de dados ou outro material, seja de forma legível por máquina ou de outra forma. Para que tais dados ou materiais sejam qualificados para a proteção de direitos autorais, o Artigo prescreve que "a seleção ou disposição de seu conteúdo [deve] constituir criações intelectuais [...]" (WTO, 1994). O Artigo estipula ainda o escopo dos direitos em tais compilações, prevendo explicitamente que "tal proteção, que não se estenderá aos dados ou materiais em si, não prejudicará qualquer direito autoral existente nos próprios dados ou materiais" (WTO, 1994).

21. O Conselho da União Europeia desenvolveu importantes declarações sobre os direitos de banco de dados (EU, 1995; EU, 1996, Artigo 7, parágrafo 4). Por exemplo: direitos legais específicos e separados duram 15 anos, mas cada vez que um banco de dados é substancialmente modificado, um novo conjunto de direitos é criado. Um proprietário tem o direito de se opor à cópia de partes substanciais de seu banco de dados, mesmo que os dados sejam extraídos e reconstruídos de forma fragmentária. A disposição, seleção e apresentação dos dados pode ser protegida por direitos

autorais, enquanto o direito de banco de dados pode proteger todo um banco de dados.

III. VISÕES, TENDÊNCIAS E DESAFIOS

22. Uma visão da Big Data no cuidado com a saúde é uma medicina personalizada, estratificada ou chamada medicina de precisão, abrangente e baseada em evidências, que combina o melhor conhecimento científico disponível com a experiência profissional dos profissionais de saúde para o benefício do paciente individual. Ela se baseia em desenvolvimentos tecnológicos em genômica e outras técnicas 'ômicas' de alta produtividade que tornaram a análise molecular de amostras humanas ordens de magnitude mais baratas e mais eficientes. Estes dados moleculares podem agora ser complementados com dados de imagem digital que vão desde o nível microscópico até a imagem de todo o corpo e com informações ambientais e de estilo de vida coletadas de um grande número de indivíduos (por exemplo, população ou grupos de pacientes), de pesquisas ou de diferentes registros, bancos de dados e infraestruturas de pesquisa. Além disso, podem ser coletados dados sobre ambiente social, comunicação e comportamento, aproximando-nos assim de uma compreensão mais abrangente da saúde e da doença, de acordo com o conceito biopsicossocial proposto pela OMS. De acordo com esta visão, todos estes dados combinados com informações nos registros eletrônicos de saúde (RES) proporcionarão no futuro uma abordagem fundamentalmente diferente para diagnosticar e tratar pacientes de forma personalizada, ou seja, oferecer as recomendações corretas e um tratamento individualizado para uma pessoa no momento certo. Também pode promover a segurança do paciente, combinando diferentes dados de diferentes fontes a fim de analisar e eventualmente prevenir eventos adversos.

23. Uma visão relacionada é que ter todo esse conhecimento profundo de um indivíduo tornará gradualmente possível determinar sua predisposição e perfil de risco para desenvolver uma doença e fornecer conselhos oportunos e direcionados para a prevenção.

24. Além disso, é provável que no futuro os cuidados de saúde se movam cada vez mais para a coleta remota de dados para diagnóstico, monitoramento e suporte terapêutico. Os dados coletados podem então ser usados para detectar alertas precoces de doenças, como um ataque cardíaco iminente, e fazer recomendações para um comportamento adequado. Também pode contribuir para melhorar os cuidados de saúde telemedicinais para pessoas que vivem em regiões remotas, permitindo melhor acesso a cuidados de saúde de qualidade e contribuindo assim para a saúde global.

25. Os pacientes podem ter a possibilidade de maior acesso e controle sobre seus dados, por exemplo, ter seus RESs em seu smartphone. O smartphone pode na verdade se tornar um dispositivo central para coordenar os próprios cuidados de

saúde e para criar sua própria rede de saúde, promovendo assim a autonomia e a alfabetização sanitária.

26. Tendo melhor acesso às informações de saúde, bem como a perfis e recomendações individualizados - desde que sejam de qualidade aprovada - os cidadãos poderão lucrar com uma melhor compreensão de seu estado de saúde e de seu comportamento de melhoria da saúde. Entretanto, existe um desafio dependente do ser humano: o fato de que mesmo que se entendesse a informação, isso não significa necessariamente agir de acordo e, de fato, muitas vezes não é tomada nenhuma ação de promoção da saúde. O tabagismo é um exemplo notável.

27. As aplicações de saúde estão desempenhando um papel cada vez maior no cuidado com a saúde e na pesquisa. Esperam-se grandes melhorias em diversas áreas, por exemplo, no monitoramento de medidas relacionadas à saúde, mudanças positivas no comportamento de saúde e apoio diagnóstico. Entretanto, até o momento, sua qualidade não é oficialmente avaliada e monitorada, e quase não existe uma estrutura regulatória específica ou um sistema de vigilância significativo em vigor. As atualizações frequentes são um obstáculo adicional para uma avaliação válida das aplicações de saúde.

28. Neste contexto, pelo menos quatro mudanças de paradigma nos cuidados de saúde individuais são prováveis de ocorrer: uma mudança da orientação da doença para a orientação da saúde; do foco na terapia para a prevenção; do aconselhamento de saúde para o estilo de vida; e do papel de um paciente para o papel de um usuário, cliente ou cidadão digital.

29. Para a indústria farmacêutica e de dispositivos médicos, existe a esperança de que Big Data promova a compreensão das doenças e seus mecanismos subjacentes, levando assim ao desenvolvimento de novos medicamentos, dispositivos e tratamentos direcionados. Espera-se também que Big Data ajude a projetar estudos clínicos estratificados, reduzindo assim o número de participantes e os custos, e que deem resultados mais rápidos.

30. Mas estas visões de Big Data não se referem apenas aos cuidados de saúde individuais e à pesquisa. As abordagens de Big Data também devem fornecer muitas informações novas a fim de fortalecer a base de evidências para políticas de saúde pública, por exemplo, permitindo melhores estratégias de prevenção ajustadas ao risco para grupos-alvo definidos.

31. Os reguladores podem compreender e controlar melhor os projetos de estudo e suas políticas podem se beneficiar de uma melhor farmacovigilância. Uma vez que um novo medicamento tenha entrado no mercado, Big Data permite a coleta e avaliação de dados do mundo real em um grande número de pacientes durante um longo período de tempo. Como a maioria da população mundial vive em áreas

cobertas por redes celulares móveis, o número de cidadãos que podem contribuir com dados está fadado a aumentar.

32. Além disso, Big Data pode contribuir para apoiar o aprendizado dos sistemas de saúde (IOM, 2007; IOM, 2013). As experiências cotidianas podem informar a melhor e mais eficiente maneira de administrar medidas diagnósticas, terapêuticas e preventivas, bem como moldar estruturas para o cuidado da saúde e pesquisa. Isto permite a análise de dados reais do mundo da saúde de uma forma estruturada e de qualidade controlada. Exemplos disso são o trabalho de aprendizado dos sistemas de saúde da OECD, projetos da UE como o Projeto Infraestruturas de Pesquisa Coordenadas Construindo Serviços de Ciência da Vida Duradouros (CORBEL, da sigla em inglês) (<http://www.corbel-project.eu/about-corbel.html>), e instrumentos práticos que facilitam o compartilhamento de dados como a Iniciativa de Medicamentos Inovadores (IMI, da sigla em inglês), e iniciativas da indústria (OECD, 2013b; OECD, 2015; IMI, 2014; PhRMA, 2014). Recentemente, a Agência Europeia de Medicamentos (EMA) publicou sua orientação sobre a liberação de dados de ensaios individuais anônimos com pacientes (EMA, 2016).

33. Tendo em mente a definição ampla de saúde pela OMS e reconhecendo que a maior parte do status de saúde de um indivíduo não depende apenas dos cuidados de saúde, mas de determinantes sociais, incluindo, mas não limitado a, desigualdades sociais, educação, estilo de vida e ambiente, Big Data abre o caminho para uma visão holística da saúde, reunindo diferentes tipos de dados, por exemplo, de registros, aplicações e registros de saúde.

34. Estas visões e tendências são acompanhadas de grandes desafios que são de natureza tecnológica, bem como ética, social e jurídica (ver Capítulo IV).

35. Acolhendo a possibilidade de uma visão holística sobre a saúde, a linha entre o setor de saúde e outros setores da sociedade é cada vez mais tênue. Além das fontes tradicionais de dados de saúde, tais como registros médicos e resultados de laboratório, outras fontes que não são tradicionalmente consideradas como relacionadas à saúde são utilizadas, por exemplo, redes sociais ou dados de consumidores, fontes de dados públicos de áreas não relacionadas à saúde, tais como a Receita Federal, os Departamentos de Educação ou de Serviços Sociais. Os provedores de mecanismos de busca também solicitam e coletam muitas informações sobre seus usuários, que posteriormente processam e vendem para diferentes corporações que por sua vez utilizam "estratégias de marketing personalizadas", oferecendo aos usuários diferentes promoções baseadas em seus históricos de busca ou participação em grupos on-line, por exemplo. Tal perfil de cliente também inclui questões de saúde - uma simples busca 'privada' na Internet em uma condição pessoal ou familiar torna-se informação publicamente acessível. Portanto, existem obviamente desafios complexos para a proteção e privacidade dos dados, bem como para a qualidade dos dados.

36. Os métodos de predição que usam aprendizagem de máquinas têm sido muito produtivos na medicina. Entretanto, é essencial compreender as suposições subjacentes e garantir que condições como estabilidade (ou seja, as condições em que os dados foram coletados permanecem as mesmas) sejam cumpridas para garantir a qualidade dos dados, bem como a validade e a utilidade das conclusões. Big Data permite principalmente encontrar padrões e correlações, que podem ser mal compreendidos como relações causais, de modo que consequências inadequadas e prejudiciais possam surgir.

37. Além disso, os algoritmos usados no reconhecimento de padrões e aplicações de saúde podem implicar em preconceitos desde o início, ou podem ser dificilmente compreendidos e controlados em termos de transparência técnica.

38. Um problema é a possível imprecisão e imprecisão na coleta e classificação dos dados devido à enorme quantidade, heterogeneidade, rápido acúmulo, rápido gerenciamento e diferentes contextos de dados. Pode haver erros no processo de coleta de dados pelo paciente/cidadão, pelo profissional médico (por exemplo, registro incompleto, classificação inadequada, análise imprecisa) ou pela tecnologia (bio)informática (por exemplo, algoritmos tendenciosos ou imprecisos, falta de interação entre clínicos e cientistas da computação). Imprecisão e erros na coleta e análise podem ter implicações adversas tanto para o único paciente/cidadão quanto para a sociedade (EC, 2015).

39. Portanto, as habilidades e competências apropriadas de todos os atores envolvidos também são um grande desafio. Os profissionais de saúde terão que entender, verificar e aplicar adequadamente os dados disponíveis. Pesquisadores, cidadãos e legisladores têm que entender a relevância dos diferentes contextos dos dados e as implicações dos algoritmos. Isto exige transparência e interpretabilidade dos dados e seus contextos. Isto só é possível se os algoritmos forem acessíveis, e se o design experimental e boas políticas de alocação de recursos existirem.

40. A linha entre as diferentes disciplinas profissionais também se tornarão tênues. Ao menos algumas doenças não serão mais compreendidas de acordo com o órgão afetado, mas sim de acordo com o mecanismo molecular subjacente, o padrão de mutações e variantes nos 'ômicos'. Uma abordagem multidisciplinar para o cuidado com a saúde e a pesquisa, como já está implementada na oncologia em muitos países, é necessária para otimizar o tratamento de pacientes individuais, o que também pode resultar em novas disciplinas profissionais. Além disso, a integração de disciplinas como ciência social, economia comportamental, epidemiologia etc. desde o início parece ser adequada em vista de uma compreensão abrangente da saúde, bem como do uso de dados em métodos de inferência causal.

41. O desenvolvimento de Big Data oferece muitas oportunidades para melhorar a saúde e pode significar economia de dinheiro para manter sistemas de saúde

sustentáveis. Entretanto, estes desenvolvimentos tecnológicos também apresentam alguns desafios para os profissionais de saúde. A implementação de muitos procedimentos tecnológicos pode reduzir o número de pelo menos alguns tipos de profissionais que o sistema de saúde precisa, eles provavelmente precisarão de habilidades diferentes, e isso pode causar uma lacuna entre diferentes gerações de profissionais de saúde. Portanto, é crucial adaptar o sistema educacional a este novo contexto e melhorar o treinamento contínuo dos profissionais de saúde.

42. Eventualmente, Big Data e um ambiente cheio de sensores levam os indivíduos a se tornarem cada vez mais transparentes, enquanto o ambiente tecnológico, os algoritmos, assim como as consequências da análise de dados e a ponderação subjacente dos fatores se tornam cada vez mais opacos. Isto leva a vários desafios éticos, especialmente no que se refere a autonomia, privacidade e justiça (ver Capítulo IV).

43. Em muitos países, particularmente no mundo em desenvolvimento, a infraestrutura ética e científica para pesquisa ainda está se desenvolvendo e muitos ainda estão longe de aproveitar os benefícios potenciais que o uso de Big Data promete. O acesso aos dados continua sendo um desafio e as plataformas de informação de saúde que podem suportar o armazenamento de enormes quantidades de dados e traduzir os dados existentes em ações reais ainda estão em falta. Isto exige alguma atenção ao desenvolvimento de capacidades. O fortalecimento de capacidades é uma forma importante de assegurar que os cientistas e profissionais de saúde do mundo em desenvolvimento não se tornem apenas coletores de dados em projetos colaborativos internacionais, mas participantes ativos na transformação de dados em benefícios tangíveis ligados a saúde para suas próprias populações.

44. Além disso, os cuidados de saúde avançados rumo à medicina de precisão são caros e limitados aos países mais ricos; e as abordagens de Big Data ainda não cobrem adequadamente as doenças que afetam as pessoas em países com infraestruturas de saúde menos avançadas. Para implementar a medicina de precisão orientada por Big Data, são necessárias novas soluções para o tratamento de dados em ambientes reais de cuidados de saúde. Somente então, Big Data poderia ser usada para melhorar as intervenções médicas, os serviços de saúde, as estratégias de predição e prevenção e as políticas de saúde em geral. Isto é uma preocupação em especial para países em desenvolvimento com recursos financeiros e humanos limitados, e diferentes conjuntos de prioridades de saúde.

45. É muito difícil prever quão rápido e até que ponto as tendências descritas acima se tornarão uma realidade generalizada. Apesar da visão otimista esboçada, é importante ser cauteloso e evitar o exagero do estado atual do conhecimento científico e os benefícios potenciais de Big Data e da medicina de precisão para o cuidado da saúde em todo o mundo. Por um lado, o 'hype de Big Data' pode levar a declarações exageradas e estimativas irrealistas, bem como à comercialização de produtos e

serviços não comprovados. Além disso, pode levar a prioridades desequilibradas de políticas de saúde, que podem ter efeitos especialmente prejudiciais para países em que o acesso a serviços essenciais não é garantido. Por outro lado, há o perigo de perder os benefícios de Big Data ao negligenciar seus potenciais. É muito importante lidar com as esperanças e promessas de uma forma equilibrada.

IV. ASPECTOS ÉTICOS

IV.1. Autonomia e Consentimento

46. A autonomia de um indivíduo em termos de exercício de sua autodeterminação inclui sete dimensões (Mertz et al., 2016):

- a. *Competência* do indivíduo interessado para acessar, compreender, avaliar e aplicar informações relevantes;
- b. A *informação* têm que estar disponível e ser relevante e compreensível para o ponto em questão;
- c. Tem que haver uma *escolha* entre diferentes opções, seja aquela entre fazer ou se abster de fazer algo ou aquela entre fazer coisas diferentes;
- d. Os *valores* do indivíduo, suas preferências e atitudes são levados em consideração ao decidir e agir;
- e. A *voluntariedade* é concedida para que o indivíduo possa decidir e agir sem coerção interna ou externa;
- f. A formação da *vontade* se refere à capacidade do indivíduo de escolher um objetivo e meios apropriados para alcançá-lo;
- g. *Ação* pode significar um ato consciente ou um refreamento.

47. A principal garantia, que tradicionalmente tem sido estabelecida para proteger a autonomia dos sujeitos humanos nos cuidados de saúde e na pesquisa, é o consentimento. Autonomia e responsabilidade, assim como consentimento e proteção de pessoas sem capacidade de consentir, são tratadas nos artigos 5, 6 e 7 da Declaração Universal sobre Bioética e Direitos Humanos (DUBDH). O artigo 6 estabelece que qualquer intervenção médica preventiva, diagnóstica e terapêutica, bem como pesquisa científica, só deve ser realizada "com o consentimento prévio, livre, expresso e informado da pessoa interessada" (Unesco, 2005). O Código de Nuremberg se concentra no consentimento e está na origem do conceito de que a participação na pesquisa é uma atividade voluntária. A Declaração de Helsinque também consagra o consentimento como uma garantia principal.

48. A escala e complexidade do processamento de dados pessoais na área de Big Data, torna virtualmente impossível para as pessoas acompanharem ou tomarem decisões significativas sobre todos os usos de seus dados pessoais. Existem grandes problemas em relação a todas as sete dimensões mencionadas acima. Mesmo que os indivíduos pudessem acompanhar e fossem informados, suas decisões poderiam ser distorcidas por várias dificuldades na tomada de decisões. Isto leva a uma perda

de autonomia que, em consequência, significa uma perda de controle e uma perda de liberdade das decisões que são tomadas pelo ambiente tecnológico e pelos processos automatizados. A governança tem que compensar esta perda de controle, assegurando que todas as dimensões da autonomia sejam respeitadas e apoiadas (ver Capítulo V).

49. Um outro problema do consentimento informado quando dado eletronicamente (consentimento informático) é a impossibilidade de identificar a idade do usuário, e assim adaptar as informações a serem apresentadas de acordo com sua idade e capacidade de compreensão (EC, 2012). A educação dos menores como usuários ativos destas novas tecnologias é particularmente importante.

50. No contexto de ensaios clínicos, o consentimento tem sido tradicionalmente específico, ou seja, os sujeitos são solicitados a consentir em um projeto de pesquisa específico. Em muitos desses casos, o consentimento obtido dos participantes da pesquisa não se estende ao uso de amostras ou dados fora da pesquisa original ou primária a qual eles consentiram. O consentimento válido exige que informações adequadas sejam divulgadas e que os sujeitos potenciais compreendam a natureza, riscos e benefícios potenciais da pesquisa e, assim, forneçam o consentimento livre e esclarecido de forma voluntária.

51. Na área de grandes bancos de dados como biobancos, e ainda mais na área de pesquisa de Big Data, o uso secundário potencial de dados no futuro não pode ser previsto. Também envolve inter-relações entre fontes de dados múltiplas e mutáveis, tanto médicas quanto não médicas. Isto levou a pedidos de um novo modelo de consentimento mais apropriado que permite um amplo uso de dados, respeitando ao mesmo tempo a autonomia do participante ou do paciente. O campo dos biobancos já apresentou novas sugestões para o fornecimento de consentimento, tal como o que é chamado de consentimento amplo.

52. O consentimento amplo pode cobrir uma ampla gama de atividades e, ao mesmo tempo, ser afunilado a usos limitados, embora apenas vagamente especificados, como a pesquisa das causas de doenças complexas. Assim, o consentimento amplo não é o oposto do consentimento "específico" porque, de acordo com o entendimento do IBC, ele não é um consentimento geral ou aberto, que não tem nenhuma condição associada. O consentimento amplo normalmente opera em um nível mais alto de abstração. Ele deixa intacto o modelo de consentimento livre e esclarecido essencial, mas um indivíduo simplesmente consente em uma série de possíveis pesquisas que poderiam ser feitas com suas informações em relação a uma área ou linha de investigação específica. As condições para que o consentimento amplo seja válido são a necessidade de alguma forma de governança e salvaguarda em que, por exemplo, um comitê relevante analisa uma proposta para assegurar que os direitos e interesses dos indivíduos sejam adequadamente protegidos ou que existam salvaguardas previstas por legislações.

53. Esta fórmula de consentimento amplo (informado) está sendo cada vez mais aceita (WMA, CIOMS), e é utilizada na Europa, local que os indivíduos consentem amplamente em todas as pesquisas possíveis com seus tecidos, relacionados à mesma área, linha de pesquisa ou similares. É também um modelo que está ganhando atenção nos países em desenvolvimento, por exemplo, no contexto da genômica e de biobancos, mas também há preocupações devido a algumas experiências negativas no campo dos biobancos. Com salvaguardas apropriadas, tais como controles para assegurar que os dados não serão utilizados para tomar uma decisão sobre o indivíduo ou de outra forma utilizados de maneira que tenha impacto sobre o indivíduo e/ou a comunidade, o uso do consentimento amplo é adequado para fins de pesquisa que contribuam para o bem público.

54. Outra abordagem é o consentimento *opt-out* de acordo com o qual os dados de saúde podem ser usados para fins de pesquisa, a menos que um indivíduo opte afirmativamente pelo *opt-out*. Esta abordagem requer informações adequadas de antemão, abordando as implicações potenciais de permitir a coleta de espécimes biológicos e dados pessoais. Portanto, exige que os profissionais forneçam educação e informações adequadas, embora o tempo muitas vezes seja escasso. Diferente do cenário de pesquisa, a impossibilidade virtual de optar por não participar dos bancos de dados comerciais relacionados à saúde (de operadoras de telefonia móvel e internet, aplicativos, bancos de dados comerciais e de comportamento do consumidor) é uma preocupação particular, pois não há virtualmente nenhum controle sobre tais bancos de dados.

55. De acordo com os princípios de participação e transparência, foi proposto um modelo adicional que depende da disponibilidade e acessibilidade das TICs. É chamado de "consentimento dinâmico". Com base no consentimento inicial do participante, implica uma atualização sobre o uso dos dados de forma contínua, de modo que o indivíduo possa optar por não permitir alguns usos específicos dos dados, enquanto permite o uso para outros fins. Requer o envolvimento dos poderes públicos tanto como garantia dos direitos individuais como também a promoção da participação dos indivíduos com educação e informação. Requer a capacitação de sujeitos e pacientes para que possam monitorar o uso de seus dados de saúde. O consentimento dinâmico poderia, em alguns ambientes, permitir o controle do acesso aos dados pelos indivíduos, possibilitado por mecanismos como os portais de consentimento. A participação contínua pode ter a vantagem de permitir que os participantes moldem as possibilidades de pesquisa ao longo de suas decisões sobre quais usos dos dados permitir, "votando" efetivamente nesses usos, autorizando-os. O indivíduo participa efetivamente na empreitada (Nuffield Council, 2015). De certa forma, o projeto torna-se uma empreitada compartilhada ou conjunta do paciente junto com o pesquisador, e assim o indivíduo não está sendo usado ou explorado. Este modelo de consentimento poderia funcionar bem em países desenvolvidos com tecnologias avançadas que permitem aos pacientes monitorar como seus dados estão sendo usados ao longo de um sistema de banco de dados, e pessoas que têm tempo

e vontade de dedicá-lo para acompanhar o uso feito de seus dados. Entretanto, sua implementação em ambientes menos desenvolvidos poderia enfrentar vários desafios, tais como a falta de tecnologias locais e baixos índices de alfabetização, a necessidade de dedicar o tempo das pessoas a outras atividades, ou a falta de conscientização e educação do público para que a compreensão dos pacientes no contexto dos cuidados de saúde e a participação na pesquisa pudessem ser dificultadas.

56. O consentimento é um grande desafio no que se refere às aplicações de saúde e na área de redes e mídias sociais. A informação digital é geralmente escrita em pequenos caracteres para ser exibida em smartphones, às vezes sem nenhuma alternativa real de discordância. O processo de dar consentimento informado na tela (e não no papel), em um ambiente cujo usuário digital não tem acesso à consulta, muitas vezes leva a clicar de forma imediata e rápida, sem tempo suficiente para fazer uma escolha informada e sem a possibilidade de verificar a real consciência, competência e voluntariedade. Os termos comerciais, que geralmente são muito longos e para a maioria das pessoas dificilmente compreensíveis, raramente são lidos. Neste sentido, é duvidoso que o procedimento de "um clique, aceitar tudo" seja realmente um consentimento livre e informado. Os meios eletrônicos de conduzir pesquisas e pesquisas de saúde podem, no entanto, ser eficientes e eficazes desde que haja salvaguardas implementadas para garantir que a autonomia dos participantes seja respeitada.

57. Diante dos desafios para obter um consentimento autônomo para a coleta e uso de Big Data e as limitações de todos os modelos de consentimento propostos, é importante que a autonomia seja apoiada pela educação pública que permita aos usuários, pacientes e participantes da pesquisa compreender o alcance e o impacto de possíveis usos de seus dados.

58. Embora deva ser enfatizado que a autonomia se liga diretamente à dignidade da pessoa, há também grandes interesses coletivos que apoiam a pesquisa que está sendo realizada no campo de Big Data e da saúde, a fim de promover a medicina de precisão baseada em evidências, assim como a aprendizagem de sistemas de saúde. Portanto, deve-se buscar um equilíbrio entre os interesses individuais e coletivos. Assim, pode ser discutida uma afirmação solidária de consentimento à participação em pesquisas de saúde baseadas em bancos de dados, contribuindo com dados anônimos - não pessoais - para grandes bancos de dados. De acordo com o princípio da reciprocidade solidária, um indivíduo que lucra com a pesquisa em saúde anterior pode ter um interesse que outras pessoas também podem lucrar com o progresso na pesquisa - pelo menos quando todos podem ter certeza de que o resultado é benéfico em termos de um bem público democraticamente definido, e quando o mau uso e a desanonimização ilegítima dos dados são fortemente processados e punidos. Em tal estrutura, há pelo menos dois elementos principais que poderiam ser úteis para equilibrar o interesse individual e coletivo: primeiro, a vulnerabilidade dos sujeitos dos

quais os dados são coletados; segundo, as finalidades para as quais os dados serão utilizados.

59. Caso se pretenda pesquisar fora do âmbito do amplo consentimento que foi obtido para o uso desses dados, é necessário um consentimento específico para o processamento de dados secundários. Este é um princípio essencial para garantir a confidencialidade e a privacidade dos dados. Entretanto, a análise secundária dos dados poderia ser eticamente admissível sem um novo consentimento informado para tal uso secundário, desde que todos os seguintes requisitos sejam atendidos:

- a. fundamento legal apropriado;
- b. avaliação pelo Comitê de Ética em Pesquisa (CEP);
- c. procedimentos técnicos adequados a fim de impedir que pesquisadores e terceiros acessem dados pessoais, tais como pseudo-anonimização;
- d. interesse público superior nesta pesquisa de saúde;
- b) inviabilidade da obtenção de um novo consentimento;
- c) os dados devem ter sido coletados de acordo com os requisitos éticos e legais.

IV.2. Privacidade e Confidencialidade

60. De acordo com o Artigo 9 da DUBDH, a privacidade da pessoa interessada e a confidencialidade de suas informações pessoais devem ser respeitadas (Unesco, 2005). Tais informações não devem ser usadas ou divulgadas para fins diferentes daqueles para os quais foram coletadas ou consentidas. Em tempos de Big Data, tal proteção de privacidade e confidencialidade está enfrentando vários desafios.

61. Os princípios tradicionais de proteção de dados, tais como limitação da finalidade, escassez e minimização de dados, proteção especial de dados sensíveis, processamento justo e salvaguarda dos direitos das pessoas envolvidas têm sido amplamente implementados a fim de proteger a privacidade, embora existam diferentes esquemas de proteção e conceitos subjacentes em diferentes regiões do mundo. Por outro lado, Big Data implica inerentemente em mudança e abertura de propósitos, ausência de limites de dados, processamento intransparente, pouca proteção de dados para obter o máximo de conhecimento possível, e intransparência para as pessoas envolvidas. Isto é especialmente verdadeiro com relação aos algoritmos que se tornam cada vez mais complicados e incompreensíveis em tempos de inteligência artificial e aprendizado profundo.

62. Além disso, a anonimização de dados pessoais em determinadas circunstâncias, e dependendo da possibilidade de acesso de outras fontes de dados, não oferece mais proteção suficiente de privacidade e confidencialidade. A integração de grandes quantidades de dados de diferentes tipos de fontes pode tornar possível a reidentificação do indivíduo em questão. Para que a anonimização continue sendo uma ferramenta útil para a proteção do indivíduo e, ao mesmo tempo, oferecendo a

possibilidade de utilização de Big Data para o bem público, são importantes salvaguardas que mitiguem o risco de reidentificação.

63. Além disso, cada vez mais são criados perfis, o que permite agir sobre grupos de acordo com perfis específicos (a chamada "identidade de grupo"). Assim, a privacidade do indivíduo não é mais adequadamente protegida mesmo que o nome do indivíduo não seja conhecido, mas, por exemplo, caso o endereço IP de seu smartphone ou computador seja. Assim, a privacidade de grupo é um conceito cada vez mais pertinente em aplicações de Big Data. A Assembleia Parlamentar do Conselho da Europa propôs até mesmo o reconhecimento de um novo direito de um sujeito a não ter seu perfil criado (Rathenau Instituut, 2017).

64. Esse novo direito reconhece que em tempos de Big Data, a privacidade como conceito é mais do que mera proteção de dados. Ao coletar, armazenar, processar, analisar e aplicar dados que depois são coletados, armazenados, etc., repetidamente (o chamado loop cibernético), isso abrange a intrusão na esfera privada, principalmente por meio de informações relacionadas a comércio, ofertas, anúncios e similares na tela de dispositivos eletrônicos, de acordo com o perfil pessoal de cada um, que é constantemente selecionado por algoritmos intransparentes com base em uma variedade de dados de diversas fontes. Esta intrusão toca várias dimensões da vida privada como comunicação pessoal, localização e participação em associações. Contra este pano de fundo, o direito de não ter seu perfil criado parece ainda mais importante.

65. O direito à privacidade está estreitamente ligado ao direito à liberdade com seus diversos aspectos legais e éticos como a liberdade de expressão, de associação, de localização, de movimento e espaço, de crenças, pensamentos e sentimentos, e de comportamento. Neste contexto, o IBC utiliza o termo privacidade no sentido de um direito ao respeito pela vida privada em relação àquelas áreas da vida ou àqueles dados que os indivíduos querem manter reservados para si mesmos ou, pelo menos, para alguns membros específicos de suas famílias ou relacionamentos.

66. Embora os indivíduos tenham interesses de privacidade, eles também compartilham interesses públicos no uso mais amplo de dados para pesquisa em saúde. Este interesse público mais amplo pode entrar em conflito com a privacidade individual. Mas a relação entre privacidade e interesse público não é simplesmente uma relação de oposição. Os dois estão implicados mutuamente: há interesses privados na realização de objetivos comuns e um interesse público na proteção da privacidade que encoraja a cooperação. Esta relação complexa leva a uma necessidade de conciliar a articulação entre o privado dentro do público e o público dentro do privado.

67. A partir de várias pesquisas sabemos que as pessoas estão conscientes de ter pouco controle sobre a coleta e mineração de seus dados, embora a maioria delas

queira ter um certo grau de controle. A longo prazo, isto pode levar a uma falta de confiança. A perda de confiança pode muito bem causar sérios danos a futuros esforços e projetos essenciais para promover a saúde pública. A fim de proteger a privacidade e o bem público ao mesmo tempo, é essencial que a população possa confiar no bom uso e proteção de seus dados de saúde ao longo de uma mistura de abordagens como lei, governança, vigilância pública, privacidade pelo projeto de um dispositivo ou de um aplicativo e privacidade por padrão, por exemplo, no que se refere a pré-requisitos de consentimento pré-instalados.

68. Novos modelos de participação na coleta de dados e mineração, como descrito no capítulo sobre autonomia, educação pública adequada sobre as implicações do uso de Big Data, gerenciamento eficaz da proteção de dados, responsabilidade de uso tanto para os participantes quanto para a sociedade em geral, bem como modelos inovadores de propriedade de dados e gestão fiduciária, devem ser desenvolvidos a fim de proteger a privacidade no amplo entendimento aqui delineado.

IV.3. Da propriedade à custódia e compartilhamento de benefícios

69. A propriedade dos dados pessoais e os direitos que os acompanham, entre os quais podemos destacar a restrição de acesso por terceiros, devem ser reconhecidos, sem dúvida, como sendo da pessoa/grupo do qual os dados derivam. A propriedade implica no direito dos sujeitos dos dados de rastrear e verificar a existência e manipulação de seus dados pessoais. Manter o controle do que somos, do que fazemos e do que pensamos e do que apresentamos aos outros e à sociedade em termos de informações pessoais é uma faceta fundamental da liberdade no século XXI (ver também Capítulo IV.2.).

70. Na área de Big Data, é importante distinguir entre dois conceitos diferentes ligados à ideia de propriedade, como é feito em áreas semelhantes como biomedicina: a de propriedade de dados pessoais e propriedade de resultados operacionais, propriedade como mecanismo de controle de dados e propriedade de produtos como algoritmos, drogas ou ferramentas de monitoramento, etc., e propriedade intelectual.

71. Argumenta-se que os benefícios dos dados serão aproveitados ao longo de práticas justas de compartilhamento de dados com a comunidade científica mais ampla e que este processo facilitará a descoberta científica. Há também uma crescente onda de estudos sobre as estruturas éticas apropriadas que devem orientar as melhores práticas de compartilhamento de dados em colaborações internacionais de pesquisa. Isto inclui o uso de métodos deliberativos para solicitar os pontos de vista dos principais interessados. Parece haver um consenso crescente na comunidade científica sobre a importância do compartilhamento de dados de pesquisa de forma mais ampla. Ao mesmo tempo, há sérias preocupações sobre o compartilhamento de dados, incluindo consentimento, propriedade dos dados, direitos de propriedade intelectual relacionados e questões de privacidade.

72. Neste contexto, fenômenos de Big Data como análise em tempo real, ligação e compartilhamento de grandes bancos de dados e uso de bancos de dados para diferentes propósitos em cuidados de saúde e pesquisa ao mesmo tempo e novamente para diferentes propósitos em um momento posterior, tornam cada vez mais problemático e inviável resolver os desafios éticos e legais ao longo de conceitos de propriedade. Devido ao equipamento técnico, habilidades e recursos financeiros, apenas poucas instituições podem lidar com bancos de dados muito grandes, excluindo assim outras de adquirir conhecimentos e desenvolver ferramentas benéficas para a saúde de todos. Assim, os conceitos tradicionais de propriedade parecem não mais ser uma estrutura normativa adequada, especialmente no que se refere à chamada divisão digital. O fenômeno de Big Data, ao contrário, levanta a necessidade de desenvolver uma estrutura normativa alternativa que incorpore novos conceitos para equilibrar interesses e benefícios legítimos.

73. O novo contexto pode ser visto como uma oportunidade para revisitar nossa visão tradicional e desenvolver novas formas éticas e legais para um esquema real de compartilhamento de benefícios. É necessário conciliar todos os direitos e interesses que se sobrepõem neste campo, tais como os da pessoa da qual os dados derivam, os do pesquisador, os das instituições privadas e organizações que utilizam os dados, e os da sociedade em geral que podem se beneficiar de tal uso. Isto se destina a mudar o foco da propriedade para a custódia e as responsabilidades de todas as partes interessadas, sendo a partilha de benefícios uma parte crucial. A custódia significa a responsabilidade pela segurança e bem-estar de alguém ou algo e representa valores éticos como cuidado, custódia (ética médica), proteção e confiança à custódia ou à guarda.

74. Neste contexto, Big Data pode ser enquadrada como um bem comum da humanidade. Avanços e novas oportunidades proporcionadas pela ciência e tecnologia podem ajudar a reduzir e não aprofundar as desigualdades que impedem muitos seres humanos de desfrutar do mais alto padrão de saúde alcançável, tanto a nível nacional como internacional. Isto está em harmonia com o artigo 2 da DUBDH que afirma: "(f) promover o acesso equitativo aos desenvolvimentos médicos, científicos e tecnológicos, assim como a maior difusão possível e o rápido compartilhamento de conhecimento relativo a tais desenvolvimentos e a participação nos benefícios, com particular atenção às necessidades de países em desenvolvimento; e (g) salvaguardar e promover os interesses das gerações presentes e futuras" (Unesco, 2005).

75. O desafio futuro será criar a melhor infraestrutura de dados disponível para uso por qualquer pessoa no mundo, desenvolvendo novas maneiras de obter lucros econômicos ao longo de serviços ou doações paralelas ou complementares. Há a necessidade de melhorar a solidariedade não apenas dos cidadãos, permitindo-lhes compartilhar ou doar seus dados para o bem comum, mas também para que atores privados compartilhem seu trabalho para o mesmo fim. Alguns projetos de vários

governos, ONGs e instituições privadas sob o nome comum de Open Data e Open Science são bons exemplos do futuro próximo, em que "a melhor infraestrutura de dados [poderia] ser disponibilizada para uso por qualquer pessoa no mundo, uma poderosa plataforma para experimentação, descoberta e criação de novas e melhores formas de vida" (Nielsen, 2014).

76. Os conceitos de responsabilidade e bem comum reforçam o que o IBC defendeu no Relatório sobre o Princípio do Compartilhamento de Benefícios: "por aceitar a ideia de direitos humanos universais, reconhecemos que estes avanços não podem ser um privilégio. O progresso não pode aprofundar as falhas existentes de desigualdade entre povos e países. Ao mesmo tempo, o relatório reconhece que a solidariedade ao longo da participação e não da beneficência é o vínculo de compartilhamento que precisa ser impulsionado. Esta é a única maneira de unir desenvolvimento e respeito por todos" (Unesco, 2015). Isto inclui os conceitos de capacitação e educação científica, transferência de conhecimento, acesso aberto à informação relacionada à saúde e empoderamento e participação na produção de conhecimento.

77. Vários programas foram iniciados para fortalecer as capacidades de pesquisadores e instituições de pesquisa em todo o mundo, especialmente nos países em desenvolvimento. No contexto da pesquisa em genômica e biobanking na África, por exemplo, a Iniciativa Hereditariedade Humana e Saúde na África (H3Africa) visa fortalecer a perícia da pesquisa em genômica e biobanking no continente para abordar problemas de saúde tanto em doenças transmissíveis como não transmissíveis (H3Africa Consortium, 2014). A expectativa é que esta iniciativa permita aos cientistas africanos contribuir significativamente para projetos científicos de ponta e também se tornar líderes no campo da genômica, bioinformática e ciência de dados. Em última análise, o objetivo é melhorar a saúde dos africanos e da comunidade global.

78. O desenvolvimento de uma estrutura ética para o compartilhamento de benefícios no uso de Big Data levanta a questão de como ela será aplicada em diferentes ambientes. Muito dependerá do contexto no qual os dados estão sendo utilizados (em pesquisas, cuidados de saúde ou estilos de vida), da natureza dos benefícios, e quais indivíduos ou grupos têm um interesse legítimo no uso dos dados. Esta diversidade, assim como o equilíbrio das vantagens e riscos de estruturas específicas de compartilhamento de benefícios, pode ser abordada de forma mais abrangente ao longo de um modelo de governança multinível (ver Capítulo V). Pode, por exemplo, implicar em dar uma quantidade considerável de lucro financeiro baseado em dados para o benefício do grupo que contribuiu substancialmente com seus dados para tornar possíveis *insights* e produtos.

IV.4. Justiça

IV.4.1. Digital Gap

79. Big Data precisa de ferramentas e fornecedores para acessá-la e gerenciá-la. A lacuna digital (também chamada divisão digital) constitui um dos maiores desafios no processo de democratização da informação e comunicação e, portanto, do desenvolvimento. A tecnologia digital é, de fato, um dos principais ativos para o desenvolvimento e para o combate a pobreza, de modo que é um dos objetivos das Metas de Desenvolvimento do Milênio das Nações Unidas em relação ao fortalecimento da parceria global para o desenvolvimento. Como resultado, a lacuna está diminuindo. Um dos objetivos era garantir que os benefícios das novas tecnologias, e especialmente das TICs, fossem disponibilizados para todos. Este programa realmente aumentou o nível de informação das populações, em particular as mais pobres e isoladas. Em relação ao setor de saúde, esta estratégia fortaleceu a prevenção de problemas de saúde graças a uma melhor promoção da saúde, mas também ampliou significativamente a oferta de cuidados por meio da telemedicina, em particular. Esta é uma evolução positiva em relação à conscientização e ao conhecimento das pessoas, que é a base de qualquer ação de saúde pública. Os bons desempenhos na telefonia móvel contribuíram para a implementação de iniciativas a fim de fortalecer os conhecimentos e habilidades do público e dos atores na luta contra doenças não transmissíveis. Por exemplo, o programa Seja Saudável, Seja Móvel é implementado em vários países (por exemplo, Índia, Egito, Senegal, Costa Rica, Zâmbia).

80. A cobertura da Internet aumentou, de pouco mais de 6% da população mundial em 2000 para 48% em 2017, e 53,6% dos domicílios cobertos em todo o mundo (ITU, 2017). Portanto, 3,6 bilhões de pessoas estão conectadas a uma rede global de aplicativos e conteúdos. Entretanto, em muitos países de baixa renda, a limitada largura de banda internacional e a infraestrutura nacional deficiente dificultam o fornecimento de serviços de Internet de banda larga a preços acessíveis. Estas limitações têm efeitos concretos sobre os tipos de serviços e aplicações que os usuários podem acessar. Consequentemente, os países em desenvolvimento e os países menos desenvolvidos experimentam limitações para integrar a gestão de Big Data em suas ofertas. Além disso, o preço médio dos serviços permanece relativamente alto em muitos dos países mais pobres, dado que a manutenção e atualização das ferramentas (por exemplo, o telefone celular) pode ser difícil.

81. Apesar de todos estes avanços no acesso às TICs, enormes disparidades permanecem e tendem a crescer entre os países, mas também dentro dos países. A análise da cobertura doméstica da Internet revela diferenças claras de acordo com o nível econômico do país. Assim, em 2017, a Europa tem uma cobertura de 84,2%, contra uma cobertura de 18% para a África (ITU, 2017). Menos da metade (41,3%) da população dos países em desenvolvimento utiliza a Internet, contra 81% nos países

desenvolvidos (ITU, 2017). Consequentemente, embora os problemas de isolamento e acessibilidade a instalações e profissionais de saúde tenham prioridade na maioria dos países em desenvolvimento, inovações tecnológicas como tele expertise, tele diagnóstico ou tele consulta, que constituem soluções eficazes, ainda não são suficientemente implementadas devido a falta de recursos.

82. Há também desigualdades significativas entre os países em termos de habilidades de TIC, alfabetização digital e existência de conteúdo local relevante. Portanto, mesmo que o conteúdo inovador seja oferecido por meio das TICs nos países em desenvolvimento, eles não podem ser implementados ali, pois requer infraestrutura e recursos humanos para sua adaptação e avaliação. Esta situação levanta problemas éticos como a transferência massiva de dados sem condições claras, com o risco de "pirataria de dados" ou políticas de governança inadequadas (Wyber et al., 2015). De fato, massas de informação são transferidas de países em desenvolvimento com recursos humanos e técnicos limitados, para serem processadas sem nenhuma garantia sobre a proteção dos dados transferidos.

IV.4.2. Não-discriminação

83. Em muitos países o acesso à saúde não se baseia em um modelo público, mas em modelos privados, que são administrados por seguradoras privadas com diferentes graus de envolvimento, controle e monitoramento por parte das autoridades públicas. Em consequência, os sujeitos podem ser negados ou impedidos de acessar os seguros devido a divulgação de dados de saúde. Mas se as companhias de seguros tomam conhecimento em detalhes dos riscos específicos de uma pessoa, se torna muito difícil falar de um modelo de seguro que tradicionalmente se baseia no equilíbrio dos riscos coletivos ao invés de perfis de risco individuais detalhados.

84. Usando TICs e abordagens de Big Data, modelos de seguro baseados no comportamento estão surgindo em diferentes tipos de seguro, incluindo o seguro de saúde. São oferecidas às pessoas taxas de seguro reduzidas ou outros tipos de prêmios se elas transferirem seus dados, por exemplo, sobre atividades esportivas, hábitos alimentares e afins, de forma regular ao longo de rastreamento, dispositivos vestíveis e aplicativos. Isto pode acabar discriminando três grupos de pessoas, seja na forma de não ter acesso a benefícios ou, o que é de se esperar a longo prazo, na forma de sofrer de desvantagens. Primeiramente, aquelas pessoas que não querem compartilhar seus dados com a seguradora são discriminadas, independentemente de poderem ou não cumprir as normas, como caminhar 10.000 passos por dia. Em segundo lugar, aquelas pessoas que não podem cumprir as normas por causa de uma deficiência, uma doença ou um grande contratempo são afetadas, independentemente de sua vontade de compartilhar os dados. Mas exatamente essas pessoas precisam de solidariedade e apoio social. Em terceiro lugar, aquelas pessoas que têm crenças diferentes das companhias de seguro sobre o que é saúde e uma vida saudável serão discriminadas. A autoridade para definir qual comportamento

relacionado à saúde é recompensado - por exemplo, andar 10.000 passos mas não meditar todas as manhãs - não deve ser com as companhias de seguro. É antes uma questão científica e social.

85. Há também um risco de discriminação proveniente de violações no anonimato oriundas de Big Data. A agregação de dados abre a possibilidade de reidentificação tendo em vista o cruzamento de referências com dados referentes a origem étnica, dados de localização, idade, outros metadados, registros de saúde ou dados genéticos (Choudhury et al., 2014).

86. Além disso, existem algoritmos poderosos, que podem traçar o perfil de indivíduos sem ter dados pessoais sobre eles, a fim de formar grupos de acordo com as características geográficas, socioeconômicas, étnicas ou outras. A anonimização dos dados pessoais pouco importa então se os resultados afetam os grupos aos quais eles pertencem com um risco de discriminação e estigmatização. Tais efeitos também afetam os membros da comunidade, que não deram seu consentimento e que não contribuíram com seus dados, mas que, apesar disso, são membros do grupo com perfil definido.

87. Além disso, uma perda de liberdade pode resultar de uma pressão sutil para ajustar o comportamento. As predições são construídas sobre a frequência do comportamento. A consequência é, por um lado, a estigmatização do comportamento pouco frequente (das minorias) e, por outro lado, uma pressão coerciva oculta para adotar "comportamento frequente" a fim de não ser excluído. Isto também se aplica a grupos.

88. Há outras possibilidades de estigmatização óbvia ou sutil e desvantagens estruturais, tais como acesso à pesquisa e progresso para aqueles com pouco equipamento e habilidades em abordagens de Big Data. Algoritmos podem conter preconceitos sistemáticos, mesmo aqueles que eventualmente contribuem para o racismo. Tendo isto em mente, tem que ser avaliado com a devida diligência em que áreas e de que forma Big Data pode contribuir para a discriminação. Medidas adequadas devem ser tomadas para evitar a estigmatização e a discriminação, de acordo com o Artigo 11 da DUBDH.

IV.5. Sustentabilidade em relação à energia e ao meio ambiente

89. Big Data para saúde é também uma questão ecológica tanto pelo seu custo energético quanto pelos ganhos que pode obter ao longo de um melhor uso dos recursos naturais. Dá-se ênfase aos benefícios potenciais, incluindo a economia de energia e a luta contra o problema do efeito estufa (Carbon War Room, sem data). Mas as tecnologias de informação com uso intensivo de energia são grandes consumidores e produzem gases de efeito estufa em todos os estágios de seu ciclo de vida (CNRS, sem data).

90. Há infraestruturas por trás de uma computação distribuída e remota que têm um alto consumo de energia e pegada de carbono: 10% da energia mundial, sendo a parte da saúde desconhecida, mas muito provavelmente significativa. Este custo é composto de 38% para os terminais, 14% para a transmissão e 48% para os centros de dados (Cappy, 2017). É válido notar que 80% do consumo de energia dos objetos conectados (14 bilhões no momento, 125 bilhões de unidades em 2025) não é dedicado ao seu uso, mas à manutenção de sua conectividade de rede. Segundo a Agência Internacional de Energia, o mundo dos dispositivos conectados consumiu 616 TWh em 2013 (IEA, 2014), que é o consumo da Suécia com seus 10 milhões de habitantes durante um ano. Os centros de dados reúnem milhares de máquinas funcionando constantemente. Além da energia necessária para operar o equipamento, as enormes necessidades de refrigeração devem ser acrescentadas. Em 2013, os centros de dados americanos consumiram cerca de 91 bilhões de quilowatts-hora de eletricidade - que seria eletricidade suficiente para alimentar todas as residências da cidade de Nova York por dois anos - e estão no caminho para atingir 140 bilhões de quilowatts-hora até 2020 (NRDC, 2015).

91. As melhorias podem resultar da equalização de voltagem, pesquisa em computação quântica, arquitetura de máquinas para resolver a lacuna entre gerenciamento e armazenamento de dados, mudança de voltagem em microprocessadores, bem como bioinspiração que visa melhorar a relação custo-benefício graças ao aprendizado de sistemas biológicos como o cérebro humano.

92. Há também preocupações sobre o possível impacto na saúde pela poluição associada aos resíduos criados pelos objetos conectados mas não mais usados. Em escala global, estima-se que a cada ano são produzidas 50 milhões de toneladas de lixo eletrônico. A iniciativa Pare o Problema de Lixo Eletrônico (StEP, da sigla em inglês) (<http://www.step-initiative.org/>), um esforço conjunto de organizações da ONU, grupos de base e indústria, prevê que até 2017 o volume total anual de lixo eletrônico terá aumentado em um terço, para 65,4 milhões de toneladas.

93. Há também preocupações com o impacto ecológico devido à busca e exploração de componentes essenciais para produzir cerca de 67 milhões de toneladas de novos equipamentos elétricos e eletrônicos que são colocados no mercado a cada ano. O termo "terra rara" descreve um conjunto de 17 elementos químicos com propriedades excepcionais. É graças a estas terras raras que as cores das telas dos computadores são tão brilhantes, que os telefones celulares têm telas de toque e as turbinas eólicas podem gerar eletricidade. O lado negativo é que sua extração e sua transformação poluem, produzem resíduos radioativos e distorcem a paisagem. A Indonésia e a Austrália já foram impactadas.

IV.6. Pesquisa

94. Como Big Data pode ser um poderoso gerador de hipóteses, particularmente para ensaios clínicos, um requisito ético primário é conduzir pesquisas seguindo os mais altos padrões de ética, proteção de dados e excelência. O biobanking está no centro da pesquisa biomédica e forma uma fonte essencial para a pesquisa de Big Data. Assim, já foram emitidas orientações de melhores práticas relativas à pesquisa de Big Data sobre amostras armazenadas em biobancos e dados coletados e armazenados em bancos de dados acadêmicos e são regularmente discutidos ao longo de recomendações de organizações internacionais como BBMRI-ERIC (<http://bbmri-eric.eu/>) e OECD. Avanços recentes incluem comitês envolvendo representantes de pacientes revisando a governança do biobanco, incluindo análise do uso do valor comercial potencial dos dados, e um processo de acompanhamento de informações sobre as várias pesquisas que utilizam os dados, permitindo que os pacientes sejam informados a longo prazo e dando-lhes uma possibilidade real de sair da pesquisa.

95. Numa tentativa de confirmar o uso de amostras armazenadas em biobancos da mais alta qualidade e dados associados, o conceito do BRIF (Fator de impacto de pesquisa de biorrecurso, do Inglês *Bioresource Research Impact Factor*) foi introduzido. A iniciativa BRIF destina-se a ajudar no acompanhamento do uso de dados e bancos de dados associados a amostras biológicas. Neste contexto, a diretriz CoBRA (Citação de Fontes Biológicas em Artigos de Pesquisa, do Inglês *Citing of Bioresources in Research Articles*) foi desenvolvida para orientar e padronizar a citação de fontes biológicas na literatura acadêmica. Isto também pode ser um exemplo para abordagens de Big Data.

96. Entendido em termos de "controle", a propriedade envolve o direito dos sujeitos dos dados de impedir qualquer tipo de manipulação de dados, incluindo a pressão da sociedade para restringir quaisquer usos inaceitáveis (ver Capítulo IV.3.). Para biobancos e bancos de dados, o controle é relevante, por exemplo, para considerar a permissibilidade de usar dados de pesquisa para fins comerciais, como é possível ao permitir que instituições privadas e de terceiros acessem amostras e dados biobancos. Pelo contrário, a propriedade dos dados pode ser um obstáculo à pesquisa, por exemplo, impedindo certas formas de mineração ou meta-análise de dados. As leis nacionais podem limitar tais pesquisas, particularmente se uma transferência de dados estiver envolvida.

97. O avanço para acesso aberto aos dados está em contradição com a proteção de dados sensíveis relevantes do ponto de vista médico. Neste contexto, é notável que algumas revistas científicas exigem a transferência de todos os conjuntos de dados utilizados para obter resultados para um repositório, a fim de permitir que outros cientistas reproduzam os resultados ou os utilizem em meta-análises. Mas isto também pode levar os editores ou outros proprietários deste tipo de repositórios a assumir direitos sobre os conjuntos de dados e também torna a mineração de dados

proibida ou muito cara. Várias leis nacionais estão atualmente em discussão para permitir o livre acesso aos conjuntos de dados para pesquisa.

98. Um grande desafio é educar os cidadãos para obter o melhor de informações pessoais derivadas de Big Data. Em primeiro lugar, os sujeitos que consentem na pesquisa devem ser capazes de exercer seus direitos de acesso com esforço razoável para que o direito possa ser exercido de forma significativa. Em segundo lugar, Big Data requer conhecimento científico avançado, obtido por meio da educação, a fim de evitar mal-entendidos, por exemplo, no que se refere ao resultado de uma análise.

99. Existe o risco de uma confiança excessiva no poder de Big Data, o que levou ao recente desenvolvimento de uma relação direta entre gerentes de dados e pacientes/cidadãos contornando a intervenção médica. Tal prática não existe no campo acadêmico, mas está se desenvolvendo rapidamente na pesquisa privada por parte de instituições comerciais. Intervenções e interpretações médicas são necessárias porque as previsões baseadas apenas na análise de Big Data não cobrem tudo o que precisa ser coberto. Entretanto, a interpretação do significado dos resultados dos Big Data para uma determinada pessoa continua sendo um objetivo fundamental da relação entre paciente e médico.

100. As análises de Big Data revelarão resultados com potencial relevância para a saúde de um determinado indivíduo, que às vezes são referidos como resultados acidentais ou não-solicitados. Como estas questões foram tratadas em um Relatório anterior do IBC sobre o Princípio da Não Discriminação e Não Estigmatização (Unesco, 2014), não há necessidade de mais comentários.

101. No nível da pesquisa em saúde pública, o uso de Big Data contribuirá para o desenvolvimento de políticas de saúde pública, para promover a saúde e prevenir doenças em nível coletivo. Devem ser implementadas salvaguardas apropriadas para proteger os direitos à privacidade dos indivíduos e devem aderir à boa governança sobre o uso dos dados. O custo de aquisição, manutenção e uso de Big Data na política de saúde pública também é de grande preocupação, especialmente em países em desenvolvimento com recursos escassos.

102. A pesquisa em saúde agora usa dispositivos pessoais conectados, tais como smartphones para rastreamento e monitoramento para coletar dados de saúde (mHealth). Isto pode ser uma fonte útil para dados do mundo real, mas também levanta questões éticas como o risco de exploração ou violação do direito à autonomia, uma vez que os telefones celulares podem coletar uma ampla gama e quantidade de informações pessoais de seus usuários com poucas salvaguardas para proteger a autonomia e a privacidade. O uso de tais dispositivos eletrônicos para conduzir pesquisas e pesquisas de saúde pode ainda ser eficiente e eficaz desde que sejam implementadas salvaguardas apropriadas para proteger a autonomia e a privacidade do indivíduo.

V. GOVERNANÇA

103. Os sistemas de governança para Big Data devem proteger os direitos fundamentais das pessoas das quais os dados se originam, incluindo sua liberdade de tomar decisões, e ter como objetivo manter a confiança do público. Assim, os princípios orientadores da governança devem incluir o respeito à autonomia e ao direito à informação, à voluntariedade, à privacidade e à proteção de dados, à transparência, à igualdade e à legalidade. A governança de dados deve garantir que o envolvimento, o engajamento, a participação e o compartilhamento de dados pelos cidadãos não serão objeto de exploração, manipulação e controle inadequado. O IBC baseou suas recomendações nestes princípios.

104. Além disso, os interesses e direitos individuais, como nas políticas de saúde pública, podem ser equilibrados com o que é chamado de bem público. No caso de Big Data, a maioria dos atores relevantes que coletam e utilizam dados geralmente carecem de um mandato democrático das pessoas cujas informações são incluídas (voluntariamente ou não) em bancos de dados para definir o que é o bem público de fato. Isto cria uma tensão ética que deve ser tratada por uma estrutura de governança apropriada que poderia garantir um "bem público".

105. Tendo em vista o escopo global de Big Data e da saúde, a enorme variedade - sempre nova - de atores neste campo e o rápido desenvolvimento tecnológico, é difícil elaborar regulamentações abrangentes e equilibradas. Mas para tornar possíveis os benefícios, para assegurar uma conduta responsável e para evitar danos é necessária uma governança prudente em diferentes níveis e de diferentes maneiras, orientando e vinculando as diferentes partes interessadas e instituições no nível legal, bem como outros níveis na forma de regras, regulamentações, decretos, acordos, códigos de conduta auto vinculativos, etc.

106. Na era de Big Data, tem se tornado cada vez mais difícil confiar na ideia de que a proteção de dados pode ser regulamentada apenas pela adesão ao consentimento ou com anonimização, sem outras salvaguardas (ver também capítulo sobre autonomia). Isto leva à conclusão de que é crucial desenvolver e aplicar uma estrutura de governança abrangente e multinível, que ainda não existe, para o uso responsável dos dados. Também é crucial observar, e já foi observado no parágrafo 35, que enquanto nesses tempos o indivíduo se torna cada vez mais transparente porque sabemos cada vez mais sobre esse indivíduo, os sistemas tendem a se tornar cada vez mais opacos. Isto precisa ser combatido com a criação de sistemas de governança mais transparentes. Tal sistema de governança pode seguir modelos diferentes, mas em todos os casos deve ser uma estrutura pública, ou pelo menos uma parceria público-privada. A participação séria dos pacientes e do público na criação e aplicação da governança é um pré-requisito para isso.

107. A transparência sobre os algoritmos, que no momento são em sua maioria opacos, é um elemento essencial neste contexto. As instituições devem ser transparentes sobre os algoritmos que utilizam. Com relação à tomada de decisões algorítmicas, elas devem oferecer explicações tanto sobre os procedimentos seguidos pelo algoritmo quanto sobre as decisões específicas que são tomadas. Alguns até afirmam que o código fonte do algoritmo deve ser publicado. Entretanto, os direitos de propriedade intelectual, bem como as características tecnológicas, por exemplo, dos algoritmos de autoaprendizagem, devem ser levados em consideração ao desenvolver esquemas adequados de transparência e controle.

108. O direito fundamental da proteção de dados oferece uma base conceitual e legal para garantir um uso moralmente e legalmente aceitável dos dados e para a criação de um sistema complexo de regras de governança da informação. O princípio da legalidade, um princípio-chave da proteção de dados, exige que todo processamento de dados pessoais necessite de uma base legal legítima. Sua essência é o sistema abrangente de controle e equilíbrio, regulando as atividades de processamento de dados pessoais. Ele visa complementar os direitos individuais, garantidos pelos direitos fundamentais à privacidade, com uma alocação efetiva das responsabilidades e deveres daqueles que utilizam os dados. O princípio da legalidade tem um status jurídico diferente em diferentes países.

109. Esta estrutura de governança deve abordar questões regulamentares e éticas relacionadas a cada etapa do ciclo cibernético, incluindo coleta, acesso, liberação, vinculação, análise, reutilização, e etc. dos dados. Com base nas considerações éticas e legais, os seguintes elementos são percebidos como cruciais para tal estrutura.

- a. Uma declaração clara da finalidade do banco de dados.
- b. Procedimentos para consentimento (amplo), re-contato (incluindo retorno dos resultados) e re-consentimento.
- c. Procedimentos para a discordância, como alternativa ao consentimento.
- d. Providências para assegurar os direitos de acesso, retificação e cancelamento de dados.
- e. Providências para a retirada, e uma descrição da medida em que a retirada é tecnicamente possível.
- f. Providências para a proteção da privacidade, pelo menos declarando os limites da proteção da privacidade.
- g. Políticas após a morte de um participante.
- h. Providências sobre a propriedade dos dados e dos produtos derivados dos mesmos.
- i. Providências para a coleta, armazenamento e duração do armazenamento dos dados, incluindo controle de qualidade e salvaguardas para proteger a privacidade e a confidencialidade.
- j. Providências para o acesso aos dados, incluindo providências para compartilhamento de dados, e critérios de acesso.

- k. Papel do CEP ou do CAD na tomada de decisões sobre compartilhamento de dados.
- l. Providências para o tratamento dos dados no caso de mudança de propriedade ou fechamento do banco de dados.
- m. Transparência dos algoritmos usados para o reconhecimento de padrões; providências para verificar o perfil de indivíduos ou grupos, de acordo com considerações éticas.
- n. Divulgação de interesse comercial e colaboração com partes comerciais.
- o. Providências que permitam aos participantes permanecer informados sobre o uso contínuo e novo dos dados, incluindo atividades de pesquisa.
- p. Uma política clara sobre a divulgação dos resultados agregados e individuais da pesquisa aos participantes.
- q. Providências para o envolvimento dos participantes na concepção de procedimentos de governança, particularmente no que se refere à supervisão ética e à comunicação com os provedores de dados.
- r. Providências para o compartilhamento de benefícios.
- s. Disposições relativas às populações indígenas/locais e minorias tradicionais.
- t. Crianças e adolescentes que atinjam a idade de maturidade durante o projeto de pesquisa devem ter a oportunidade de dar consentimento informado para o armazenamento e uso continuado de seus dados, e também devem poder retirar o consentimento para pesquisas futuras.

110. Estes elementos poderiam ser mais desenvolvidos em tratados internacionais, regulamentações auto vinculativas entre fornecedores ou mesmo levar à criação de uma agência global de vigilância. As entidades existentes, tais como Comitês de Ética em Pesquisa (CEPs), Comitês de Acesso aos Dados (CADs) e agências de proteção de dados poderiam desempenhar um papel importante em tal estrutura de governança multinível, abordando as relações entre as múltiplas partes interessadas.

111. O IBC recomenda que um Instrumento Jurídico Internacional sobre Proteção de Dados seja adotado pelos Estados-Membros. Se solicitado, o IBC está disposto a contribuir para o desenvolvimento de uma estrutura de governança adequada que deverá preceder tal tratado. Cada país pode então criar dentro de sua própria jurisdição legislação específica, sob a qual pode ser criada uma agência encarregada de supervisionar esses sistemas de governança. Tal agência também proporcionaria um ponto de entrada claro para o controle público.

VI. RECOMENDAÇÕES

112. A fim de fazer dos Big Data um sucesso global para a saúde, para colher as oportunidades de Big Data na área de saúde e pesquisa e, ao mesmo tempo, evitar a violação dos direitos humanos fundamentais consagrados na Declaração Universal dos Direitos Humanos e na Declaração Universal sobre Bioética e Direitos Humanos,

o IBC dá as seguintes recomendações.

113. Diante da natureza complexa, do escopo global e da grande variedade de partes interessadas envolvidas em Big Data relacionada à saúde, a cooperação internacional, bem como uma abordagem de governança multinível, são cruciais para que a confiança e o controle sejam equilibrados para o benefício de todos (para mais detalhes ver Capítulo V).

114. O IBC considera quatro medidas cruciais para proteger os direitos individuais e promover o bem público, ao mesmo tempo em que reconhece a inevitável perda de controle pelos indivíduos sobre o uso de seus dados em tempos de Big Data: governança, educação, capacitação e compartilhamento de benefícios.

115. A seguir, o IBC apresenta exemplos de medidas importantes. Eles são entendidos como um passo em um debate global contínuo. Eles se dirigem a diferentes partes interessadas. O sucesso final dessas medidas dependerá da cooperação e participação abrangente de todas as partes interessadas, incluindo pacientes, sujeitos de pesquisa, usuários e cidadãos em geral.

116. As agências internacionais são chamadas a desenvolver e apoiar uma estrutura global para o uso de Big Data em áreas relacionadas à saúde, especialmente em cuidados de saúde e pesquisa. Por exemplo:

- a. A ONU é recomendada a desenvolver e adotar um Instrumento Legal Internacional sobre Proteção de Dados em Saúde e Pesquisa em Saúde.
- b. A Unesco é recomendada a desenvolver uma convenção sobre a proteção da privacidade, incluindo uma estrutura para novas abordagens de propriedade e custódia de dados pessoais, incluindo mas não se referindo apenas a dados de saúde. Ela também deve abordar o tratamento de dados pessoais e a presença digital de um indivíduo após sua morte física. Esta convenção pode se basear na proposta de resolução do Conselho de Direitos Humanos sobre o direito à privacidade na era digital (A/HRC/34/L.7/Rev.1).
- c. Recomenda-se que a OMS encoraje o estabelecimento de um acordo por lojas de aplicativos sobre a apresentação de aplicações relacionadas à saúde de forma que a autonomia, transparência e informação adequada sejam garantidas.
- d. Um sistema de vigilância global para o uso de Big Data em aplicações relacionadas à saúde deve ser estabelecido.
- e. Os acordos internacionais devem adotar um entendimento de Big Data como um bem comum da humanidade e facilitar o acesso aberto e o uso de Big Data para o bem comum, sempre que possível. Para este fim, deve ser criada uma infraestrutura de dados públicos.
- f. Recomenda-se que a OECD desenvolva uma estrutura para compartilhar os benefícios das aplicações de Big Data. Isto também deve contribuir para

- superar cada vez mais a lacuna digital.
- g. Instituições internacionais, bem como instituições regionais e nacionais que estabelecem normas técnicas para dispositivos e processos no contexto de Big Data são chamadas a definir essas normas de acordo com os princípios éticos de respeito à autonomia, privacidade e justiça, conforme elaborados neste relatório, bem como com a necessidade de transparência.
 - h. A Agência Internacional de Energia (AIE) é chamada a sinergizar esforços para um uso sustentável e responsável da energia no gerenciamento de Big Data. Ela poderia criar programas para desenvolver políticas de eficiência energética, incluindo o diálogo com países não-membros.
 - i. Agências de proteção ambiental, como a ONU Meio Ambiente, juntamente com a OMS, são chamadas a estabelecer um plano de ação coordenado para economizar recursos raros. Além disso, elas deveriam implementar um programa para evitar que os resíduos eletrônicos tenham impacto sobre a saúde das pessoas, especialmente nos países em desenvolvimento.

117. Os governos nacionais são chamados a desenvolver e lançar um plano de ação que inclua legislação e políticas que devem, entre outros, abordar os seguintes aspectos:

- a. Implementação internacionalmente harmonizada dos princípios de proteção de dados aceitos globalmente.
- b. Os governos nacionais - na medida em que ainda não o tenham feito - são chamados a estabelecer uma agência de controle de dados eficaz. Todas as agências nacionais de controle de dados devem trabalhar em conjunto com o Sistema de Vigilância Global mencionado acima (parágrafo 116[d]) para o uso de Big Data em aplicações relacionadas à saúde. Este sistema coordenado poderia servir como ponto de partida para o Instrumento Jurídico Internacional sobre Proteção de Dados em Saúde e Pesquisa em Saúde.
- c. Capacitação para cuidados de saúde e pesquisas relacionadas a Big Data, incluindo uma infraestrutura de dados eficiente.
- d. Promoção da aprendizagem de sistemas de saúde que utilizam dados dos cuidados diários de saúde para produzir evidências para os cuidados de saúde futuros.
- e. Possibilitar uma cooperação transfronteiriça eficaz no processamento de dados pessoais de saúde para fins de interesse público relacionado à saúde.
- f. Promoção da educação em relação às habilidades, competências e consciência de Big Data sobre as implicações eticamente relevantes. Especial atenção deve ser dada a grupos vulneráveis como menores e pessoas com deficiências.
- g. Introdução de diversos modelos de consentimento que são especificados de acordo com o contexto da saúde e da pesquisa, permitindo um consentimento amplo e dinâmico quando apropriado, viável e

- salvaguardado. O mesmo se aplica ao consentimento eletrônico.
- h. Proteção da privacidade dos indivíduos que compartilham seus dados por padrão, bem como pelo projeto do dispositivo e do equipamento tecnológico. Deve ser dada atenção separada à proteção da privacidade de grupo - não menos importante, tendo como pano de fundo as múltiplas possibilidades de discriminação.
 - i. Implementação de um Comitê de Ética em Pesquisa ou instituição similar que supervisione a pesquisa comercial de Big Data.
 - j. Desenvolvimento e implementação de instrumentos para um envolvimento significativo do público e do paciente de uma forma sensível à cultura.
 - k. Os governos devem lançar programas coordenados que motivem os desenvolvedores de dispositivos a criar protocolos de comunicação que permitam a economia de energia. Uma prioridade deve ser a padronização de procedimentos para coleta e gerenciamento de dados.

118. A grande variedade de instituições de saúde e pesquisa, assim como corporações, são chamadas a desenvolver diretrizes sensíveis ao contexto, código de conduta e instrumentos auto vinculantes com foco especial em aplicações de Big Data. Políticas éticas e códigos de conduta também devem ser desenvolvidos para profissionais de saúde, cientistas da computação, pesquisadores clínicos, cientistas de dados e outras partes interessadas. A motivação do comportamento ético deve ser o objetivo principal, mas também devem ser previstas consequências para o caso de violação dos princípios éticos.

BIBLIOGRAFIA

Nota do tradutor: As citações e referências bibliográficas foram mantidas no idioma e formato originais.

- African Union (AU). 2014. African Union Convention on Cybersecurity and Personal Data Protection. Addis Ababa, AU. Available at: <https://www.au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection>
- Andanda, P., Wathuta, J., Leisinger, K. and Schroeder, D. 2016. National and International Compliance Tools, a report for the TRUST Project. Available at: <http://trust-project.eu/wp-content/uploads/2017/02/TRUST-664771-National-and-International-Compliance-Tools-Final.pdf>
- Asian-Pacific Economic Cooperation (APEC). 2015. Updates to the APEC Privacy Framework. Singapore, APEC Secretariat. Available at: http://mddb.apec.org/Documents/2016/SOM/CSOM/16_csom_012app17.pdf
- Bowker, G. C. 2014. Big data, big questions the theory/data thing, International Journal of Communication. Los Angeles, USC Annenberg Press, Vol. 8, pp. 1795-1799.
- Cappy, A. 2017. Impact énergétique des Big Data (Energy Impact of Big Data). In: Bouzeghoub, M. and Mosseri, R. eds. Les Big Data à découvert (Big Data Uncovered). Paris, CNRS.
- Carbon War Room. n.d. Carbon War Room (online). Available at: <http://carbonwarroom.com/> Centre national de la recherche scientifique (CNRS). n.d. EcolInfo (online). Paris, CNRS. Available at: <http://ecoinfo.cnrs.fr/>
- Choudhury, S., Fishman, J. R., McGowan, M. L., and Juengst, E. T. 2014. Big Data, open science and the brain: lessons learned from genomics, Frontiers in Human Neuroscience. Lausanne, Frontiers, 8, 239. doi:10.3389/fnhum.2014.00239
- Cohen, J. E. 2013. What privacy is for, Harvard Law Rev. Cambridge, Harvard Law Review, Vol. 126, No. 7, pp. 1904-1933.
- Coles, D., Wathuta, J. and Andanda, P. 2016. ICT and Mobile Data for Health Research. In: Schroeder, D., Cook Lucas, J., Fenet, S. and Hirsch, F. eds. "Ethics Dumping" – Paradigmatic Case Studies, a report for the TRUST Project. Available at: <http://trust-project.eu/wp-content/uploads/2016/03/TRUST-664771-Paradigmatic-Case-Studies-WP1-Final.pdf>
- Comitato Nazionale per la Bioetica (CNB). 2016. Information and Communication Technologies and Big Data: Bioethical Issues. Rome, CNB. Available at: http://bioetica.governo.it/media/172149/p124_2016_information-technologies-and-big-data_en.pdf
- Coordinated Research Infrastructures Building Enduring Life-science Services (CORBEL). n.d. Available at: <http://www.corbel-project.eu/home.html>
- Council of Europe (CoE). 1950. Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights). Strasbourg, CoE. Available at: <http://www.echr.coe.int/pages/home.aspx?p=basictexts>

- CoE. 1981. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Strasbourg, CoE. Available at: <http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108>
- CoE. 2001. Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, regarding supervisory authorities and transborder data flows. Strasbourg, CoE. Available at: <http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/181>
- European Commission (EC). 2012. Opinion No. 26: Ethics of Information and Communication Technologies, European Group on Ethics in Science and New Technologies (EGE). Luxembourg, Publications Office of the European Union. Available at: <http://bookshop.europa.eu/en/ethics-of-information-and-communication-technologies-pbNNAJ12026/>
- European Commission (EC). 2015. Opinion No. 29: The ethical implications of new health technologies and citizen participation, European Group on Ethics in Science and New Technologies (EGE). Luxembourg, Publications Office of the European Union. Available at: <https://publications.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-8529-01aa75ed71a1/language-en/format-PDF/source-39541144>
- European Court of Justice (ECJ). 2012. Judgement of the Court (Grand Chamber): Used Soft GmbH v Oracle International Corp (C-128/11). Luxembourg, ECJ. Available at: <http://curia.europa.eu/juris/document/document.jsf?docid=124564&doclang=EN>
- European Medicines Agency (EMA). 2016. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. London, EMA. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2016/03/WC500202621.pdf
- European Union (EU). 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Union (online). Available at: <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046>
- EU. 1996. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Official Journal of the European Union (online). Available at: <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A31996L0009>
- EU. 2015. General Data Protection Regulation (compromise text after trilogue) (online). Available at: <http://www.statewatch.org/news/2015/dec/eu-council-dp-reg-draft-final-compromise-15039-15.pdf>
- EU. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union (online). Available at: <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679>

- H3Africa Consortium. 2014. Research capacity. Enabling the genomic revolution in Africa, *Science*. New York, American Association for the Advancement of Science (AAAS), Vol. 344, No. 6190, pp. 1346-1348.
- Hoeren, T. 2014. Big Data and the Ownership in Data: Recent Developments in Europe, *European Intellectual Property Review*. London, Sweet & Maxwell, Vol. 36, No. 12, pp. 751- 754.
- Innovative Medicines Initiative (IMI). 2014. Code of Practice on Secondary Use of Medical Data in Scientific Research Projects, final draft 27 August 2014. Available at: [http://www.imi.europa.eu/sites/default/files/uploads/documents/CodeofPractice_SecondaryUseDRAFT .pdf](http://www.imi.europa.eu/sites/default/files/uploads/documents/CodeofPractice_SecondaryUseDRAFT.pdf)
- Institute of Medicine (IOM). 2007. *The Learning Healthcare System: Workshop Summary*. Washington D.C., The National Academies Press.
- IOM. 2013. *Best care at lower cost: The path to continuously learning health care in America*. Washington D.C., The National Academies Press.
- International Data Corporation (IDC). 2014. *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives* (online). Available at: <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>
- International Energy Agency (IEA). 2014. *More Data, Less Energy: Making Network Standby More Efficient in Billions of Connected Devices*. Paris, IEA. Available at: https://www.iea.org/publications/freepublications/publication/MoreData_LessEnergy.pdf
- International Telecommunication Union (ITU). 2015a. *ICT Facts and Figures 2015*. Geneva, ITU. Available at: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>
- ITU. 2015b. *ITU's Annual Report 2015*. Geneva, ITU. Available at: <http://www.itu.int/en/annual-report-2015/Pages/default.aspx>
- ITU. 2016. *ITU's Annual Report 2016*. Geneva, ITU. Available at: <http://www.itu.int/en/annual-report-2016/Pages/default.aspx>
- ITU. 2016. *ICT Facts and Figures 2016*. Geneva, ITU. Available at: <http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>
- ITU. 2017. *ICT Facts and Figures 2017*. Geneva, ITU. Available at: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>
- Kaplan, W., Wirtz, V. J., Mantel-Teeuwisse, A., Stokl, P., Duthey, B. and Laing, R. 2013. *Priority Medicines for Europe and the World: 2013 Update*. Geneva, WHO. Available at: [http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf? ua=1](http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf?ua=1)
- Lavenex S. and Schimmelfennig F. 2009. EU rules beyond EU borders: theorizing external governance in European politics, *Journal of European public policy*. Taylor and Francis Online, Vol. 16, No. 6, pp. 791-812.
- Laurie, G., Ainsworth, J., Cunningham, J., Dobbs, C., Jones, K. H., Kalra, D., Lea, N. C. and Sethi, N. 2015. On moving targets and magic bullets: Can the UK lead the way

with responsible data linkage for health research?, *International Journal of Medical Informatics*. Amsterdam, Elsevier, Vol. 84, pp. 933-940.

Lea, N. C. 2015. Design and Development of a Knowledge Modelling Approach to Govern the Use of Electronic Health Records for Research. Doctoral thesis, University of London, UK.

Mertz M., Jannes, M., Schlomann, A., Manderscheid, E., Rietz, C. and Woopen, C. 2016. Digitale Selbstbestimmung. Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health. DOI 10.8716/ceres/00001,

Mittelstadt, B. D. and Floridi L. 2016. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts, *Sci Eng Ethics*. New York, Springer, Vol. 22, pp. 303-341.

Mostert, M., Bredenoord, A. L., Biesart, M. C. I. H. and van Delden, J. J. M. 2015. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach, *European Journal of Human Genetics*. London, Nature Publishing Group, Vol. 24, No. 7, pp. 956-960. doi:10.1038/ejhg.2015.239

Natural Resources Defense Council (NRDC). 2015. America's Data Centers Consuming and Wasting Growing Amounts of Energy (online). New York, NRDC. Available at: <https://www.nrdc.org/resources/americas-data-centers-consuming-and-wasting-growing-amounts-energy>

Nielsen, M. 2014. Who Owns Big Data? (online). BBVA OpenMind. Available at: <https://www.bbvaopenmind.com/wp-content/uploads/2014/02/BBVA-OpenMind-Technology-Innovation-Internet-Business-Michael-Nielsen-Who-Owns-Big-Data.pdf>

Nuffield Council on Bioethics. 2015. The collection, linking and use of data in biomedical research and health care: ethical issues. London, Nuffield Council on Bioethics. Available at: http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf

Organisation for Economic Cooperation and Development (OECD). 2010. Improving Health Sector Efficiency: The Role of Information and Communication Technologies. Paris, OECD. Available at: <http://www.oecd.org/els/health-systems/improvinghealthsectorefficiency.htm>

OECD. 2013a. Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data. Paris, OECD. Available at: <https://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf>

OECD. 2013b. Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Protection Challenges. Paris, OECD. Available at: <http://www.oecd.org/publications/strengthening-health-information-infrastructure-for-health-care-quality-governance-9789264193505-en.htm>

OECD. 2015. Health Data Governance: Privacy, Monitoring and Research. Paris, OECD. Available at: <http://www.oecd.org/fr/publications/health-data-governance-9789264244566-en.htm>

OECD. 2017. OECD Recommendation on Health Data Governance. Paris, OECD. Available at: <http://www.oecd.org/els/health-systems/health-data-governance.htm>

Pharmaceutical Research and Manufacturers of America (PhRMA)/European Federation of Pharmaceutical Industries and Associations (efpia). 2013. Principles for

- Responsible Clinical Trial Data Sharing (online). Available at: <http://transparency.efpia.eu/uploads/Modules/Documents/data-sharing-prin-final.pdf>
- Rathenau Instituut. 2017. Human rights in the robot age. Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality. The Hague, Rathenau Instituut. Available at: <https://www.rathenau.nl/en/publication/human-rights-robot-age-challenges-arising-use-robotics-artificial-intelligence-and>.
- Science International. n.d. Open Data in a Big Data World. Paris, International Council for Science (ICSU), International Social Science Council (ISSC), the World Academy of Sciences (TWAS) and the InterAcademy Partnership (IAP). Available at: <http://www.science-international.org/>
- Sethi, N. and Laurie, G. 2013. Delivering proportionate governance in the era of eHealth: making linkage and privacy work together, *Med Law Int.* Thousand Oaks, SAGE Publications, Vol. 13, No. 2-3, pp. 168–204.
- Shilton, K. 2009. Four billion little brothers? Privacy, mobile phones, and ubiquitous data collection, *Communications of the ACM.* New York, ACM, Vol. 52, No. 11, pp. 48-53.
- United Nations (UN). 1948. Universal Declaration of Human Rights. New York, UN. Available at: <http://www.un.org/en/universal-declaration-human-rights/index.html>
- UN. 1990. Guidelines for the Regulation of Computerized Personal Data Files, adopted by the General Assembly resolution 45/95 of 14 December 1990. New York, UN. Available at: <http://www.refworld.org/pdfid/3ddcafaac.pdf>
- United Nations Conference on Trade and Development (UNCTAD). 2016. Data protection regulations and international data flows: Implications for trade and development. Geneva, UNCTAD. Available at: http://unctad.org/en/PublicationsLibrary/dtlstict2016d1_en.pdf
- United Nations Educational, Scientific and Cultural Organization (Unesco). 2005. Universal Declaration on Bioethics and Human Rights. Paris, Unesco. Available at: http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html
- Unesco. 2014. Report of the International Bioethics Committee of Unesco (IBC) on the Principle of Non-discrimination and Non-stigmatization. Paris, Unesco. Available at: <http://unesdoc.unesco.org/images/0022/002211/221196E.pdf>
- Unesco. 2015. Report of the International Bioethics Committee of Unesco (IBC) on the Principle of the Sharing of Benefits. Paris, Unesco. Available at: <http://unesdoc.unesco.org/images/0023/002332/233230E.pdf>
- United States of America (USA). 1996. Health Insurance Portability and Accountability Act. Washington D.C., US Congress. Available at: <https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf>
- USA. 2002. HIPAA Privacy Rule. Washington D.C., US Department of Health and Human Services. Available at: <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html?language=es>
- USA. 2009. The American Recovery and Reinvestment Act. Washington D.C., US Congress. Available at: <https://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf>

- USA. 2014. Big Data: Seizing Opportunities, Preserving Values. Washington D.C., Executive Office of the President. Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.1_4_final_print.pdf
- USA. n.d. Privacy Shield Program Overview (online). Washington D.C., US Department of Commerce. Available at: <https://www.privacyshield.gov/Program-Overview>
- Vayena, E. and Tasioulas, J. 2015. "We the Scientists": a Human Right to Citizen Science. *Philos. Technol.* New York, Springer, Vol. 28, pp. 479-485
- Weiss, M. A. and Archick, K. 2016. U.S.-EU Data Privacy: From Safe Harbor to Privacy Shield. Washington D.C., US Congressional Research Service. Available at: <https://fas.org/sgp/crs/misc/R44257.pdf>
- World Health Organization (WHO). 2011. mHealth: New horizons for health through mobile technologies: second global survey on eHealth, Global Observatory for eHealth series-Volume 3. Geneva, WHO. Available at: http://apps.who.int/iris/bitstream/10665/44607/1/9789241564250_eng.pdf
- WHO. n.d. eHealth (online). Available at: <http://www.who.int/ehealth/en/>
- World Medical Association (WMA). 2013. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects (online). Available at: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
- WMA. 2016. WMA Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks (online). Available at: <https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/>
- World Trade Organization (WTO). 1994. Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). Geneva, WTO. Available at: https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm
- Wyber, R., Vaillancourt, S., Perry, W., Mannava, P., Folaranmi, T. and Celi, L.A. 2015. Big data in global health: improving health in low- and middle-income countries, *Bulletin of the World Health Organization*. Geneva, WHO, Vol. 93, pp. 203-208. doi: <http://dx.doi.org/10.2471/BLT.14.139022>

ANEXO B – LISTA DOS ARTIGOS SELECIONADOS

Após a busca inicial de artigos e verificação de critérios de inclusão e exclusão, detalhadas na metodologia, definiu-se o *corpus* da pesquisa, sendo este composto por 44 artigos, além do Relatório do IBC sobre Big Data e Saúde (4).

A lista com os artigos selecionados pode ser vista abaixo.

Allen J, Adams C, Flack F. The role of data custodians in establishing and maintaining social licence for health research. *Bioethics* 2019; 33(4):502-510.

Altomare DF. Defining authorship in the era of Big Data collection and its consequences on the academic career. *International Journal of Colorectal Disease* 2019; 34(5):939-940.

Arga KY. COVID-19 and the futures of machine learning. *OMICS: a journal of integrative biology* 2020; 24(9):512-514.

Barocas S, Boyd D, Friedler S, Wallach H. Social and Technical Trade-Offs in Data Science. *Big Data* 2017; 5(2):71-72.

Bassan S, Harel O. The ethics in synthetics: statistics in the service of ethics and law in health related research in Big Data from multiple sources. *Cleveland-Marshall Journal of Law and Health* 2018; 31(1):87-117.

Beier K, Schweda M, Schicktanz S. Taking patient involvement seriously: a critical ethical analysis of participatory approaches in data-intensive medical research. *BMC Medical Informatics and Decision Making* 2019; **19**:90.

Bernstein J. Not the Last Word: Big Data Will Make You Confront Big Ethical Questions — Here's Why. *Clinical Orthopaedics & Related Research* 2019; 477(2):284-287.

Canales C, Lee C, Cannesson M. Science without conscience is but the ruin of the soul: the ethics of Big Data and artificial intelligence in perioperative medicine. *Anesthesia and Analgesia* 2020; 130(5):1234-1243.

- Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of Big Data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Medicine* 2019; **17**:143–147.
- Cassel C, Bindman A. Risk, benefit, and fairness in a Big Data world. *The Journal of the American Medical Association* 2019; **322**(2):105-106.
- Cesare N, Lee H; McCormick T, Spiro E, Zagheni E. Promises and pitfalls of using digital traces for demographic research. *Demography* 2018; **55**(5):1979-1999.
- Demotes-Mainard J, Cornu C, Guérin A, Bertoye PH, Boidin R, Bureau S, et al. How the new european data protection regulation affects clinical research and recommendations? *Therapies* 2019; **74**(1):31-42.
- Dhar V. Should we regulate digital platforms? *Big Data* 2017; **5**(4):277-278.
- Dorey CM, Baumann H, Biller-Andorno N. Patient data and patient rights: swiss healthcare stakeholders' ethical awareness regarding large patient data sets – a qualitative study. *BMC Medical Ethics* 2018; **19**(1):1-14.
- Dove ES, Özdemir V. What role for law, human rights, and bioethics in an age of Big Data, consortia science, and consortia ethics? The importance of trustworthiness. *Laws* 2015; **4**(3):515-540.
- Evans BJ. Power to the people: data citizens in the age of precision medicine. *Vanderbilt Journal of Entertainment and Technology Law* 2017; **19**(2):243-265.
- Fisher CB, Layman DM. Genomics, Big Data, and Broad Consent: a New Ethics Frontier for Prevention Science. *Prevention Science* 2018; **19**(7):871-879.
- French R. What is the good of it — ethical controls of human subject health research? *Journal of Bioethical Inquiry* 2018; **15**(4):589-602.
- Goldenholz DM, Goldenholz SR, Krishnamurthy KB, Halamka J, Karp B, Tyburski M, et al. Using mobile location data in biomedical research while preserving privacy. *Journal of the American Medical Informatics Association* 2018; **25**(10):1402-1406.
- Goodman KW. Ethics in health informatics. *Yearbook of Medical Informatics* 2020; **29**(1):26-31.

- Gürsoy G, Harmanci A, Tang H, Ayday E, Brenner SE. When Biology Gets Personal: Hidden Challenges of Privacy and Ethics in Biological Big Data. Pacific Symposium on Biocomputing 2019, p. 386-390.
- Hallowell N, Parker M, Nellåker C. Big data phenotyping in rare diseases: some ethical issues. *Genetics in Medicine* 2019; 21(2):272-274.
- Ienca M, Ignatiadis K. Artificial intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB Neuroscience* 2020; 11(2):77-87.
- Lajaunie C, Ho CWL. Pathogens collections, biobanks and related-data in a one health legal and ethical perspective. *Parasitology* 2018; 145(5):688-696.
- Laviolle B, Denèfle P, Gueyffier F, Bégué É, Bilbault P, Espérou H, et al. The contribution of genomics in the medicine of tomorrow, clinical applications and issues. *Therapies* 2019; 74(1):9-15.
- Lawler M, Maughan T. From Rosalind Franklin to Barack Obama: data sharing challenges and solutions in genomics and personalised medicine. *The New Bioethics* 2017; 23(1):64-73.
- Lawler M, Morris AD, Sullivan R, Birney E, Middleton A, Makaroff L, et al. A roadmap for restoring trust in Big Data. *The Lancet Oncology* 2018; 19(8):1014-1015.
- Manrique de Lara A, Peláez-Ballestas I. Big data and data processing in rheumatology: bioethical perspectives. *Clinical Rheumatology* 2020; 39(4):1007-1014.
- Margetts H, Dorobantu C. Rethink government with AI. *Nature* 2019; 568:163–5.
- Martín-Ruiz ML, Fernández-Aller C, Portillo E, Malagón J, Barrio C del. Developing a System for Processing Health Data of Children Using Digitalized Toys: Ethical and Privacy Concerns for the Internet of Things Paradigm. *Science and Engineering Ethics* 2018; 24(4):1057-1076.
- Martins TGS, Costa ALFA, Martins TGS. Big data use in medical research. *Einstein (São Paulo)* 2018; 16(3):1-2.
- Middleton A. Society and personal genome data. *Human Molecular Genetics* 2018; 27(1):8-13.
- Mittelstadt BD, Floridi L. The ethics of Big Data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics* 2016; 22:303-341.

- Peirce AG, Elie S, George A, Gold M, O'Hara K, Rose-Facey W. Knowledge development, technology and questions of nursing ethics. *Nursing Ethics* 2020; 27(1):77-87.
- Resnick KS, Appelbaum PS. Passive Monitoring of Mental Health Status in the Criminal Forensic Population. *The Journal of the American Academy of Psychiatry and the Law* 2019; 47(4):456-466.
- Robinson K. A false promise of COVID-19 'big' health data? Health data integrity and the ethics and realities of Australia's health information management practice. *Health Information Management Journal* 2020; 50(1-2):9-12.
- Rosen DL, Buchbinder M, Juengst E, Rennie S. Public health research, practice, and ethics for justice-involved persons in the Big Data era. *American Journal of Public Health* 2020; 110(S1):S37-38.
- Rothstein MA. Ethical issues in Big Data health research: currents in contemporary bioethics. *Journal of Law, Medicine & Ethics* 2015; 43(2):425-429.
- Spector-Bagdady K, Jagsi R. Big data, ethics, and regulations: Implications for consent in the learning health system. *Medical Physics* 2018; 45(10):e845–847.
- Stoeklé HC, Mamzer-Bruneel MF, Frouart CH, Le Tourneau C, Laurent-Puig P, Vogt G, et al. Molecular Tumor Boards: Ethical Issues in the New Era of Data Medicine. *Science and Engineering Ethics* 2018; 24(1):307-322.
- Tossas-Milligan KY, Winn RA. Breaking the cycle of health inequities: the bioethics of data. *Journal of Health Care for the Poor and Underserved* 2019; 30(5):86-90.
- Überall M, Werner-Felmayer G. Integrative biology and big-data-centrism: mapping out a bioscience ethics perspective with a S.W.O.T. Matrix. *OMICS: A Journal of Integrative Biology* 2019; 23(8):371-379.
- Vezyridis P, Timmons S. Resisting Big Data exploitations in public healthcare: free riding or distributive justice? *Sociology of Health & Illness* 2019; 41(8):1585-1599.
- Wang S, Bonomi L, Dai W, Chen F, Cheung C, Bloss CS, et al. Big Data Privacy in Biomedical Research. *IEEE Transactions on Big Data* 2020; 6(2):296-308.

ANEXO C – RELATÓRIO DE ANÁLISE DE DADOS DO IRAMUTEQ

Durante a pesquisa, foi utilizado como ferramenta de apoio à análise o software IRaMuTeQ. Dentre os recursos que o IRaMuTeQ oferece, foram utilizados os recursos de Análise de Classificação Hierárquica Descendente e de Análise de Especificidades.

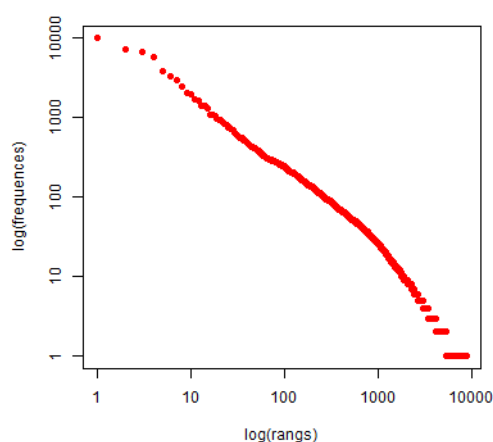
A seguir, são apresentados os dados provenientes das análises acima mencionadas, além de alguns dados estatísticos sobre o *corpus* que o IRaMuTeQ fornece ao realizar análises neste.

- Dados estatísticos

O *corpus* geral foi constituído por 44 textos, separados em 4996 segmentos de texto (ST), com aproveitamento de 4497 STs (90,01%). Emergiram 180990 ocorrências (palavras, formas ou vocábulos), sendo 8799 palavras distintas e 3479 com uma única ocorrência.

Ainda, conforme é possível ver na Figura 11, o software também disponibiliza o diagrama de Zipf, sendo este um gráfico que apresenta a distribuição de frequência das palavras no *corpus* (44).

Figura 11 – Gráfico de Zipf.



Fonte: Exportado da análise do IRaMuTeQ pelo autor.

- Análise de Classificação Hierárquica Descendente

Na análise de classificação, o software identificou quatro classes de palavras. O conteúdo analisado foi categorizado em quatro classes:

- Classe 1, com 1418 STs (31,53% do *corpus*);
- Classe 2, com 1182 STs (26,28% do *corpus*);
- Classe 3, com 529 STs (11,76% do *corpus*); e,
- Classe 4, com 1368 STs (30,42% do *corpus*).

Ainda, os resultados que apontam a presença de determinada forma em cada classe são exibidos dentro da interface do próprio IRaMuTeQ, de modo que capturas de tela deste são apresentadas a seguir, por conterem os dados brutos acerca das diferentes formas identificadas dentro de cada classe.

Figura 12 – Formas presentes na Classe 1.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme		p
0		137	178	76.97	177.21	nom	ai		< 0,0001
1		108	129	83.72	167.55	nom	algorithm		< 0,0001
2		75	83	90.36	135.56	nr	machine_learning		< 0,0001
3		75	90	83.33	114.15	ver	monitor		< 0,0001
4		85	108	78.7	114.05	ver	bias		< 0,0001
5		95	130	73.08	107.02	nom	model		< 0,0001
6		99	138	71.74	106.61	nom	disease		< 0,0001
7		89	122	72.95	99.65	nom	population		< 0,0001
8		54	60	90.0	96.29	nr	mental		< 0,0001
9		59	72	81.94	86.13	ver	sequence		< 0,0001
10		100	153	65.36	83.95	nom	digital		< 0,0001
11		44	48	91.67	81.27	nom	pattern		< 0,0001
12		41	45	91.11	74.74	nom	brain		< 0,0001
13		36	37	97.3	74.74	ver	predict		< 0,0001
14		48	58	82.76	71.42	nr	facebook		< 0,0001
15		81	122	66.39	70.59	nom	treatment		< 0,0001
16		35	37	94.59	68.72	nom	diagnosis		< 0,0001
17		33	36	91.67	60.79	nom	estimate		< 0,0001
18		30	32	93.75	57.79	nom	prediction		< 0,0001
19		26	26	100.0	56.78	nom	phone		< 0,0001
20		31	34	91.18	56.45	nom	diagnostic		< 0,0001
21		25	25	100.0	54.59	nom	forensic		< 0,0001
22		58	85	68.24	54.06	ver	test		< 0,0001
23		54	77	70.13	54.06	ver	trace		< 0,0001
24		28	30	93.33	53.43	nom	volume		< 0,0001
25		27	29	93.1	51.25	nom	neuroscience		< 0,0001

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 13 – Formas presentes na Classe 2.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p	
0		286	434	65.9	389.06	nr	ethical	< 0,0001	
1		117	196	59.69	118.06	nr	healthcare	< 0,0001	
2		55	69	79.71	103.23	nom	stakeholder	< 0,0001	
3		41	45	91.11	98.59	nr	participatory	< 0,0001	
4		119	221	53.85	91.13	nom	issue	< 0,0001	
5		73	116	62.93	82.53	nom	participation	< 0,0001	
6		43	56	76.79	74.64	nom	perspective	< 0,0001	
7		185	423	43.74	73.39	nr	big_data	< 0,0001	
8		66	110	60.0	66.15	nom	principle	< 0,0001	
9		247	626	39.46	65.13	adj	patient	< 0,0001	
10		36	48	75.0	59.43	nom	moral	< 0,0001	
11		114	242	47.11	57.24	nom	medical	< 0,0001	
12		30	40	75.0	49.43	nom	intensive	< 0,0001	
13		85	172	49.42	49.4	ver	practice	< 0,0001	
14		54	94	57.45	48.12	nom	registry	< 0,0001	
15		36	55	65.45	44.09	nom	literature	< 0,0001	
16		21	25	84.0	43.22	nom	tension	< 0,0001	
17		74	150	49.33	42.55	ver	question	< 0,0001	
18		25	33	75.76	42.0	nom	interviewee	< 0,0001	
19		32	48	66.67	40.83	nr	normative	< 0,0001	
20		88	192	45.83	39.56	ver	concern	< 0,0001	
21		43	74	58.11	39.33	ver	discuss	< 0,0001	
22		100	230	43.48	36.99	ver	benefit	< 0,0001	
23		23	31	74.19	36.98	nom	involvement	< 0,0001	
24		40	69	57.97	36.31	nom	think	< 0,0001	
25		332	685	33.71	35.85	nom	research	< 0,0001	

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 14 – Formas presentes na Classe 3.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p	
0		146	240	60.83	588.14	nom	science	< 0,0001	
1		69	73	94.52	489.62	nom	consortium	< 0,0001	
2		41	65	63.08	167.31	adj	big	< 0,0001	
3		35	57	61.4	137.06	nom	industry	< 0,0001	
4		33	55	60.0	124.82	nr	global	< 0,0001	
5		28	43	65.12	119.06	nom	actor	< 0,0001	
6		24	33	72.73	119.04	nom	profit	< 0,0001	
7		41	85	48.24	111.03	nr	bioethics	< 0,0001	
8		13	13	100.0	97.79	nr	bioeconomy	< 0,0001	
9		73	239	30.54	85.77	nom	ethic	< 0,0001	
10		41	99	41.41	85.74	nom	law	< 0,0001	
11		36	81	44.44	84.88	nom	company	< 0,0001	
12		21	34	61.76	82.52	nom	era	< 0,0001	
13		12	13	92.31	81.49	nom	alliance	< 0,0001	
14		12	14	85.71	73.99	nr	centrism	< 0,0001	
15		19	31	61.29	73.77	ver	encourage	< 0,0001	
16		15	21	71.43	72.36	adj	extreme	< 0,0001	
17		10	11	90.91	66.55	nr	ga4gh	< 0,0001	
18		16	25	64.0	66.09	nom	trustworthiness	< 0,0001	
19		28	65	43.08	62.3	nom	society	< 0,0001	
20		22	45	48.89	60.36	nom	innovation	< 0,0001	
21		9	10	90.0	59.1	nr	fundes	< 0,0001	
22		28	67	41.79	59.08	adj	open	< 0,0001	
23		29	75	38.67	53.19	nom	commercial	< 0,0001	
24		65	248	26.21	52.77	nom	share	< 0,0001	
25		7	7	100.0	52.59	nom	commercialization	< 0,0001	

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 15 – Formas presentes na Classe 4.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42% *			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p	
0		308	379	81.27	505.53	ver	consent	< 0,0001	
1		171	197	86.8	309.42	nom	protection	< 0,0001	
2		127	153	83.01	206.93	nom	regulation	< 0,0001	
3		310	556	55.76	192.39	nom	information	< 0,0001	
4		173	249	69.48	189.98	nom	subject	< 0,0001	
5		155	215	72.09	185.26	ver	inform	< 0,0001	
6		775	1872	41.4	182.64	nom	datum	< 0,0001	
7		145	219	66.21	139.32	nom	participant	< 0,0001	
8		186	317	58.68	128.63	nom	privacy	< 0,0001	
9		67	76	88.16	121.76	nom	custodian	< 0,0001	
10		118	173	68.21	121.38	nom	personal	< 0,0001	
11		49	52	94.23	101.2	nr	identifiable	< 0,0001	
12		66	82	80.49	98.92	nom	requirement	< 0,0001	
13		83	116	71.55	95.17	ver	protect	< 0,0001	
14		103	160	64.38	90.37	nom	purpose	< 0,0001	
15		94	142	66.2	88.67	ver	review	< 0,0001	
16		418	985	42.44	86.04	nom	research	< 0,0001	
17		58	73	79.45	84.28	nr	institutional	< 0,0001	
18		37	38	97.37	81.15	nr	hipaa	< 0,0001	
19		147	270	54.44	78.33	ver	access	< 0,0001	
20		198	398	49.75	77.07	nom	individual	< 0,0001	
21		52	65	80.0	76.59	nom	board	< 0,0001	
22		61	83	73.49	74.12	nom	security	< 0,0001	
23		52	66	78.79	74.03	nom	secondary	< 0,0001	
24		47	59	79.66	68.48	nom	release	< 0,0001	
25		70	104	67.31	68.44	nom	rule	< 0,0001	

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

- **Análise de Especificidades**

Na análise de especificidades, os resultados que apontam a predominância de uma variável são exibidos dentro da interface do próprio IRaMuTeQ, de modo que capturas de tela deste são apresentadas a seguir, por conterem os dados brutos acerca das variáveis identificadas como predominantes em cada classe.

Na parte da tela marcada em vermelho é possível observar cada variável (identificada por um asterisco *) presente na classe. Observa-se três tipos de variáveis: as relativas à área do primeiro autor, as relativas aos países dos artigos e as relativas aos artigos presentes no *corpus*, tendo sido essa última utilizada para controle interno do Autor durante a utilização do software.

Figura 16 – Variáveis identificadas na Classe 1.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme		p
781		748	1457	51.34	391.63		*Saude		< 0,0001
782		135	158	85.44	220.45		*43_Passive_Monitorin...		< 0,0001
783		112	124	90.32	204.14		*51_Artificial_Intellige...		< 0,0001
784		161	212	75.94	203.26		*33_Promises_and_Pitf...		< 0,0001
785		682	1651	41.31	115.49		*Estados_Unidos		< 0,0001
786		81	110	73.64	92.59		*41_Science_Without_...		< 0,0001
787		359	835	42.99	62.4		*Tecnologia		< 0,0001
788		35	40	87.5	58.56		*Turquia		< 0,0001
789		35	40	87.5	58.56		*55_COVID_19_and_th...		< 0,0001
790		66	106	62.26	47.49		*45_The_contribution_...		< 0,0001
791		37	58	63.79	28.33		*39_Not_the_Last_Word		< 0,0001
792		25	34	73.53	27.99		*04_Breaking_the_Cycl...		< 0,0001
793		65	125	52.0	24.95		*31_Knowledge_devel...		< 0,0001
794		13	16	81.25	18.38		*32_Big_Data_use_in_...		< 0,0001
795		13	16	81.25	18.38		*Brasil		< 0,0001
796		31	53	58.49	18.05		*50_Rethink_governm...		< 0,0001
797		65	135	48.15	17.8		*48_Molecular_Tumor...		< 0,0001
798		18	26	69.23	17.21		*24_When_Biology_Ge...		< 0,0001
799		25	41	60.98	16.61		*49_Big_data_phenoty...		< 0,0001
800		64	140	45.71	13.46		*19_Integrative_Biolog...		0.00024
801		64	140	45.71	13.46		*Austria		0.00024
802		122	305	40.0	10.87		*Suica		0.00097
803		30	63	47.62	7.66		*21_Beyond_the_hype...		0.00564
804		30	63	47.62	7.66		*Singapura		0.00564
805		26	59	44.07	4.35		*08_A_false_promise_o...		0.03697
806		153	430	34.85	2.48		*Franca		NS (0.11504)

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 17 – Variáveis identificadas na Classe 2.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme		p
562		4	8	50.0	2.33	sw	try		NS (0.12719)
563		4	8	50.0	2.33	num	31		NS (0.12719)
564		59	191	30.89	2.18	sw	into		NS (0.13945)
565		5	11	45.45	2.09	num	36		NS (0.14811)
566		885	3296	26.85	2.04	sw	and		NS (0.15276)
567		11	29	37.93	2.04	sw	despite		NS (0.15286)
568		126	181	69.61	182.74		*25_Patient_data_and_...		< 0,0001
569		116	164	70.73	173.55		*11_Taking_patient_inv...		< 0,0001
570		116	164	70.73	173.55		*Alemanha		< 0,0001
571		276	564	48.94	170.78		*Etica		< 0,0001
572		120	207	57.97	112.44		*17_What_Is_the_Good...		< 0,0001
573		134	305	43.93	52.61		*Suica		< 0,0001
574		162	414	39.13	38.84		*Australia		< 0,0001
575		66	134	49.25	37.61		*Mexico		< 0,0001
576		66	134	49.25	37.61		*02_Big_data_and_dat...		< 0,0001
577		125	324	38.58	27.25		*Bioetica		< 0,0001
578		51	114	44.74	20.56		*Administracao		< 0,0001
579		37	76	48.68	20.02		*30_Resisting_big_data...		< 0,0001
580		15	23	65.22	18.09		*28_Social_and_Techni...		< 0,0001
581		113	336	33.63	10.12		*01_The_Ethics_of_Big...		0.00147
582		21	50	42.0	6.45		*36_Pathogens_collect...		0.01112
583		44	125	35.2	5.27		*31_Knowledge_devel...		0.02163
584		14	32	43.75	5.07		*40_Public_Health_Res...		0.02428
585		14	38	36.84	2.2		*53_Should_We_Regul...		NS (0.13758)
586		37	115	32.17	2.11		*12_Ethics_in_Health_I...		NS (0.14606)

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 18 – Variáveis identificadas na Classe 3.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p	
457		2	6	33.33	2.69	sw	wonder	NS (0.10078)	
458		2	6	33.33	2.69	num	62	NS (0.10078)	
459		2	6	33.33	2.69	sw	ourselves	NS (0.10078)	
460		2	6	33.33	2.69	num	52	NS (0.10078)	
461		2	6	33.33	2.69	num	43	NS (0.10078)	
462		12	66	18.18	2.66	sw	now	NS (0.10299)	
463		6	28	21.43	2.54	sw	either	NS (0.11129)	
464		4	17	23.53	2.28	sw	neither	NS (0.13139)	
465		37	253	14.62	2.11	sw	there	NS (0.14593)	
466		9	49	18.37	2.08	sw	indeed	NS (0.14909)	
467		7	36	19.44	2.06	sw	around	NS (0.15093)	
468		3	12	25.0	2.03	sw	until	NS (0.15411)	
469		198	274	72.26	1028.9		*05_What_Role_for_Law	< 0,0001	
470		300	949	31.61	456.56		*Reino_Unido	< 0,0001	
471		238	654	36.39	447.21		*Direito	< 0,0001	
472		24	66	36.36	39.05		*10_From_Rosalind_Fr...	< 0,0001	
473		24	84	28.57	23.3		*26_Society_and_perso...	< 0,0001	
474		9	19	47.37	23.3		*44_A_roadmap_for_	< 0,0001	
475		22	76	28.95	21.99		*30_Resisting_big_data...	< 0,0001	
476		27	114	23.68	16.01		*Administracao	< 0,0001	
477		25	140	17.86	5.17		*19_Integrative_Biolog...	0.02298	
478		25	140	17.86	5.17		*Austria	0.02298	
479		21	115	18.26	4.8		*12_Ethics_in_Health_I...	0.02845	
480		29	182	15.93	3.18		*13_Power_to_the_Peo...	NS (0.07460)	
481		10	53	18.87	2.61		*50_Rethink_governm...	NS (0.10632)	

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor

Figura 19 – Variáveis identificadas na Classe 4.

1418/4497 31.53%		1182/4497 26.28%		529/4497 11.76%		1368/4497 30.42%			
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p	
596		3	5	60.0	2.07	num	2002	NS (0.15030)	
597		3	5	60.0	2.07	sw	k	NS (0.15030)	
598		12	28	42.86	2.06	sw	said	NS (0.15131)	
599		121	148	81.76	190.55		*29_How_the_new_Eur...	< 0,0001	
600		108	133	81.2	166.98		*18_Genomics_Big_Da...	< 0,0001	
601		104	155	67.1	102.02		*07_The_Ethics_in_Syn...	< 0,0001	
602		97	148	65.54	89.18		*37_The_role_of_data_...	< 0,0001	
603		40	48	83.33	64.18		*46_Using_mobile_loc...	< 0,0001	
604		92	155	59.35	63.5		*Espanha	< 0,0001	
605		92	155	59.35	63.5		*42_Developing_a_Syst...	< 0,0001	
606		187	414	45.17	46.86		*Australia	< 0,0001	
607		196	439	44.65	46.52		*Franca	< 0,0001	
608		91	182	50.0	34.35		*13_Power_to_the_Peo...	< 0,0001	
609		140	336	41.67	21.7		*01_The_Ethics_of_Big...	< 0,0001	
610		40	75	53.33	18.92		*06_Ethical_Issues_in_...	< 0,0001	
611		30	59	50.85	11.79		*08_A_false_promise_o...	0.00059	
612		233	654	35.63	9.8		*Direito	0.00174	
613		29	60	48.33	9.22		*03_Big_Data_Privacy_i...	0.00239	
614		22	44	50.0	8.05		*20_Big_data_ethics_a...	0.00455	
615		23	50	46.0	5.8		*36_Pathogens_collect...	0.01604	
616		13	27	48.15	4.03		*Italia	0.04461	
617		13	27	48.15	4.03		*23_Defining_authors...	0.04461	
618		11	23	47.83	3.31		*22_Risk_Benefit_and_...	NS (0.06890)	
619		528	1651	31.98	3.0		*Estados_Unidos	NS (0.08322)	
620		273	835	32.69	2.51		*Tecnologia	NS (0.11342)	

Fonte: Captura de tela do software IRaMuTeQ realizada pelo Autor