

UMA CONJECTURA SOBRE AS TECNOLOGIAS DE BIG DATA NA PRÁTICA JURÍDICA

A CONJECTURE ABOUT BIG DATA TECHNOLOGIES IN LEGAL PRACTICE.

MÁRCIO PUGLIESI*

ANDRÉ MARTINS BRANDÃO**

RESUMO

As tecnologias da informação e comunicação passaram por uma rápida evolução nas últimas décadas. A possibilidade de criar conteúdo foi majorada a níveis nunca antes vistos. As tecnologias de big data trabalham com a problemática de navegar de modo mais eficiente nesse oceano de dados recriado diariamente. Foram desenvolvidas com o intuito de buscar por correlações nesses dados, com a finalidade de extrair informações úteis. A lógica indutiva é utilizada para dar vazão a eficiência e rapidez, em contraposição a clássica lógica dedutiva. Nesse contexto, põe-se a seguinte questão: como as tecnologias de big data podem transformar a prática jurídica na sociedade contemporânea? Tem-se como hipótese que o uso das tecnologias de big data gera impactos sobre os mais diversos campos que são tocados pelo fenômeno jurídico, como privacidade, discriminação e tomada de decisão. Busca-se, a partir dessa

ABSTRACT

Information and communication technologies have undergone rapid development in recent decades. In contemporary society, the ability to create content was increased to levels never seen before. Big data technologies work with the possibility to navigate more efficiently in this data ocean recreated daily. These technologies were developed to search for correlations in data sets, in order to extract useful information. Inductive logic is used to enable efficiency and speed, in opposition to classical deductive logic. In this context, a question appears: how big data technologies may transform legal practice in contemporary society? This work has for hypothesis that the use of big data technologies generates impacts on various fields touched by the legal phenomenon, such as privacy, discrimination and automation of decision-making. This work seeks to build a conjecture about possible changes in legal practice from these impacts.

* Doutor e Livre-Docente em Direito pela Universidade de São Paulo. Doutor em Filosofia pela Pontifícia Universidade Católica de São Paulo. Professor da Pontifícia Universidade Católica de São Paulo no Programa de Pós-Graduação.

Email: mpugliesi@hotmail.com

** Aluno do Curso de Doutorado de Filosofia do Direito da Pontifícia Universidade Católica de São Paulo. Bolsista CAPES. Mestre em Direito pela Universidade da Amazônia - UNAMA. Pesquisador do grupo de estudos sobre direito, análise, informação e sistemas (GEDAIS) da PUC/SP.

Email: andrembrandao@gmail.com

problemática, construir uma conjectura sobre possíveis transformações na prática jurídica a partir desses impactos. Foi observado que há possibilidade de uso das tecnologias de *big data* de forma que auxilie na tomada de decisão, assim como é possível o contrário. É necessário o conhecimento profundo de seu funcionamento e estrutura, em ordem de possibilitar seu uso de modo a auxiliar na tomada humana de decisão. Por outro lado, essas tecnologias podem criar novos tipos de discriminação, além de ameaçarem cada vez mais a privacidade e até o processo pessoal de construção de identidade. Existem paradoxos na tecnologia, que podem sofrer menos tensão a partir da construção de uma ética do big data.

PALAVRAS-CHAVE: Big data. Prática jurídica, Lógica indutiva. Reconhecimento de padrões. Tomada de decisão. Privacidade.

On one hand, there are possibilities for the use of big data technologies in order to assist in the legal decision-making process, as well as the reverse is possible. It is necessary to understand the structure and functions of big data technologies deeply, in order to enable its use to help human decision-making. On the other hand, these technologies may create new forms of discrimination, and also increasingly threaten privacy and even the personal identity construction process. There are paradoxes in these technologies that may suffer less stress with the construction of a big data ethics.

KEYWORDS: Big data. Legal practice. Inductive logic. Pattern recognition. Decision-making. Privacy.

1. INTRODUÇÃO

As tecnologias da informação e comunicação (TIC) passaram por uma evolução que beira a ficção científica nas últimas décadas. Seja pela velocidade na comunicação e no acesso à informação, pelos os novos hábitos consequentes das redes sociais ou pelo constante controle de informação – os tempos estão mudando, diria Dylan.

Nesse contexto, a presente conjectura põe a seguinte questão: como as tecnologias de big data e analytics podem transformar a prática jurídica na sociedade de controle e informação? Tem-se como hipótese que o uso das tecnologias de big data gera impactos sobre os mais diversos campos que são tocados pelo fenômeno jurídico, como privacidade, discriminação, direito público à informação e até automação da tomada de decisão (jurídica), dentre muitos outros. Nesse sentido, faz-se necessário uma conjectura para compreender as transformações por qual passa o fenômeno jurídica na atual sociedade com o advento dessas tecnologias.

Para responder o questionamento primeiramente é necessário investigar as tecnologias de big data. Parte-se do pressuposto que a possibilidade de criar conteúdo foi majorada a níveis nunca antes vistos. A tecnologias de big data trabalham com o seguinte

questionamento: de que modo poderia se navegar de modo mais eficiente nesse oceano de dados recriado diariamente? Foram criadas com o intuito de buscar por correlações nesse oceano de dados, com a finalidade de extrair informações úteis. A lógica indutiva é utilizada para dar vazão a eficiência e rapidez, em contraposição a clássica lógica dedutiva.

A partir da investigação do sentido das tecnologias de big data, poder-se-á passar à problemática central do trabalho, ou seja, de que forma as ferramentas de big data podem transformar a prática jurídica. Tem-se por hipótese, como visto, que a utilização de tais tecnologias cria impactos em campos tocados pelo direito. Busca-se, a partir dessa problemática, construir uma conjectura sobre possíveis transformações na prática jurídica a partir desses impactos.

2. AS TECNOLOGIAS DE BIG DATA

Com a rápida evolução das tecnologias da informação e comunicação (TIC) nos últimos dez anos a possibilidade de criar conteúdo foi amplificada de forma nunca antes vista. É estimado que entre a invenção da escrita e 2006 a humanidade acumulou cerca de 180 Exabytes de dados. Esse número cresceu dez vezes entre 2006 e 2011, chegando a um total de 1600 Exabytes (FLORIDI, 2012).

O aumento exponencial do volume de dados é acelerado pelas aplicações de redes sociais, que permitem a qualquer sujeito criação de grande quantidade de conteúdo facilmente e sem custos. Os smartphones ainda aceleram mais esse processo – o sujeito passa a dispor de uma porta de entrada nômade, capaz de gerar conteúdo em qualquer lugar. A internet das coisas (*internet of things* – IOT) também é um fator que deve começar a ser considerado, multiplicando as fontes artificiais e autônomas de dados (que não deixam de ser sujeitos, nessa conjectura). Casas inteligentes, onde dispositivos são interligados em sistemas, também começam a se tornar realidade e geram quantidades de dados sem precedentes. Tem-se a expectativa de extrair informações valiosas dessa multiplicidade com novas tecnologias (ERMOLAYEV V. et al, 2014, p. 4).

“As a result, there is much talk about Big Data.”¹ (FLORIDI, 2012, p. 436) O termo Big Data é definido, pela ASE International Conference (*Academy of Science and Engineering, de Harvard*), como: “(...) *large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.*”² (2014, p. 01).

O aumento do poder de processamento por meio de núcleos múltiplos, o aumento da velocidade das redes de computador e a computação em paralelo, dentre outros, permitem hoje aos sujeitos, inclusive privados, a exploração do Big Data. O artigo sobre a tecnologia de *Map-Reduce* de J. Dean e S. Ghemawat (2008) foi importante por levar a diversos desenvolvimentos em Big Data. Apesar da tecnologia de *Map-Reduce* já ser conhecida à época, a junção com o *Apache Hadoop* permitiu o processamento de grandes conjuntos de dados por qualquer pessoa em um cluster de nodos. Como já visto, a geração de dados nunca foi mais rápida, e o que faz o Big Data uma ferramenta importante hoje é exatamente a acessibilidade – a possibilidade de lidar com essa tamanha quantidade.

Nunca antes foi necessário lidar com tamanha quantidade de dados – trata-se de um problema epistemológico do Big Data – como conhecer (lidar) com essa complexidade? Outro problema gerado é de ordem ética: como usar esses dados? (FLORIDI, 2012).

A aproximação da problemática epistemológica mostra que o Big Data lida com o oceano de complexidade de forma a reconhecer pequenos padrões ou correlações. Com a rapidez e facilidade de geração de dados, por meio de redes sociais, por exemplo, trabalha-se com a possibilidade de achar onde novos padrões se encontram e trabalhar com eles, para os diferentes fins que se almeja – os quais entram no campo da problemática ética. Assim, o ponto nevrálgico do Big Data é reconhecimento de padrões

1 “Como resultado, muito se fala sobre o *big data*.”

2 “Largo, diverso, complexo, longitudinal, e/ou distribuído conjunto de dados gerados por instrumentos, sensores, transações na internet, e-mail, video, *click streams*, e/ou qualquer outra fonte digital disponível hoje e no futuro.”

ou correlações em complexidades de dados, gerando informações e conhecimentos úteis e relevantes (FLORIDI, 2012).

Diversos profissionais na área da informação utilizam Big Data com a finalidade de coleta e análise de dados. “*Fundamentally, big data analytics is a workflow that distills terabytes of low-value data down to, in some cases, a single bit of high-value data. The goal is to see the big picture from the minutia of our digital lives*”³ (FISHER et al. 2012, p. 53).

Alguns teóricos chegam a falar não mais de uma sociedade de informação ou economia do conhecimento, mas de uma economia de dados e sociedade de dados: “*It’s all data now: Data Economy and Data Society. This is a confession that we are no longer in control of the knowledge contained in the data our systems collect.*”⁴ (GRELLER, 2012).

Big Data, de certa forma, é um termo utilizado para descrever “*large volumes of unstructured and structured content—usually in amounts measured in terabytes or petabytes—that enterprises want to harness and analyze.*”⁵ As tecnologias tradicionais de gerenciamento de bases de dados, de caráter relacional, tem dificuldade em lidar com a grande quantidade de inserção de dados, tornando difícil manipulá-los. Desse modo, são necessárias novas tecnologias para lidar com essas dificuldades (ERMOLAYEV V. et al, p. 32).

A análise da massiva quantidade de dados criada e recriada a todo o momento requer ferramentas diversas da ciência tradicional, principalmente pela necessidade de velocidade no processamento. A lógica indutiva aparece como a saída para analisar a grande quantidade de dados que compõe o sistema de big data de modo veloz,

3 “Fundamentalmente, *big data analytics* é um fluxo de trabalho que destila *terabytes* de dados de baixo valor para, em alguns casos, um único *bit* de dados de alto valor. O objetivo é ver o quadro geral a partir dos pequenos eventos de nossas vidas digitais.”

4 “Tudo é dados agora: Economia de dados e sociedade de dados. Essa é uma confissão que não estamos mais no controle do conhecimento contido nos dados coletados por nossos sistemas.”

5 “Grandes volumes de conteúdo estruturado e não estruturado – geralmente em quantidades medidas em *terabytes* ou *petabytes* – que empresas querem coletar e analisar.”

tendo por característica essencial o potencial de reconhecimento de correlações ou padrões e de previsibilidade.

Um estudo mais profundo das relações entre lógica indutiva e dedutiva é necessário para apreender tais conceitos de modo mais completo. Bryan Skyrms afirma que “a lógica está associada ao raciocínio correto e à formulação de argumentos convincentes. Muitos ignoram que a lógica tem dois ramos principais: a lógica indutiva e a lógica dedutiva.” (SKYRMS, 1966, p. 11). A primeira delas está ligada fortemente à noção de probabilidade: “*Die Wahrscheinlichkeitstheorie enthält das Induktionsproblem, und das Wahrscheinlichkeitsproblem kann nicht ohne eine Antwort auf die Frage der Induktion gelöst werden.*”⁶ (REICHENBACH, p. 212, 1983).

O argumento, parte central da lógica, “é uma coleção de sentenças em que uma é indicada como conclusão e as demais indicadas como premissas (do argumento em pauta).” Os enunciados, como átomos dos argumentos, são sentenças que englobam “uma asserção fatural definida.” Nesse sentido, “argumento é uma coleção de enunciados, (...) a conclusão enuncia o ponto a defender e as premissas enunciam as razões oferecidas em favor da conclusão.” (SKYRMS, 1966, p. 12-13). A lógica tem por função estudar a força de evidência que liga premissas e conclusões em argumentos.

Em certos casos a verdade das premissas garante a verdade da conclusão – o liame entre elas é o mais forte possível. Em outros, as premissas apresentam algum apoio para a conclusão, mas não garantem integralmente sua correção - essa ligação, por sua vez, não é tão forte. “Os argumentos podem ter vários graus de força.” (SKYRMS, 1966, p. 16-17).

Os primeiros tipos de argumento, nos quais as premissas asseguram a verdade da conclusão, são chamados dedutivamente válidos. Enquanto os outros, nos quais as premissas oferecem certa evidência para a conclusão, mas não a asseguram completamente, são indutivamente fortes. “Um argumento é dedutivamente válido

6 “A teoria da probabilidade envolve o problema da indução, e o problema da probabilidade não podem ser resolvidos sem ter uma resposta para a questão da indução.”

se e somente se é impossível que sua conclusão seja falsa quando suas premissas são verdadeiras.” No segundo caso, “um argumento é indutivamente forte se e somente se é improvável que sua conclusão seja falsa quando suas premissas são verdadeiras, e o argumento, além disso, não é dedutivamente válido.” (SKYRMS, 1966, p. 18).

No argumento dedutivamente válido as asserções de ordem fatural são encerradas nas premissas, sendo logicamente impossível uma conclusão falsa quando as premissas são verdadeiras. A conclusão se encontra dentro do espaço das premissas, não o extravasando. De modo diferente, em um argumento indutivamente forte a conclusão extravasa as asserções contidas nas premissas, podendo ocorrer situações onde as premissas são verdadeiras, e falsa a conclusão. Apesar do risco de erro dos argumentos indutivos, trazem uma vantagem sobre os dedutivamente válidos – a possibilidade de previsão e descoberta de fatos novos, baseando-se em alguns conhecidos.

A força indutiva dos argumentos é medida por um tipo de probabilidade – a probabilidade indutiva. “*Inductive logic is here conceived as the theory of an explicatum for probability*”⁷ (CARNAP, 1963, p. 161). A probabilidade indutiva não depende somente de premissas ou conclusões consideradas isoladamente, mas da correlação entre premissas e conclusões, que se mostra em termos de evidência. Trata-se de um modo de reconhecimento de padrões ou correlações.

A lógica pode ser definida como o estudo da força da relação de evidência entre premissas e conclusões de argumentos. Há pelo menos duas formas de avaliação dessa força – validade dedutiva e força indutiva. A lógica dedutiva realiza testes de validade dedutiva – estuda regras para verificar quando um argumento é válido; enquanto a lógica indutiva mede a probabilidade indutiva de argumentos – sua força indutiva, e como elaborar argumentos indutivamente fortes (CARNAP, 1963).

Os argumentos se diferenciam principalmente por sua força de evidência – como em uma escala decrescente: começando pelos dedutivamente válidos e indo até os destituídos de valor. “A

7 “Lógica indutiva é aqui concebida como a teoria de uma explicação para a probabilidade.”

lógica dedutiva e a indutiva não se distingue em função dos tipos de argumentos de que tratam, mas em função de padrões diversos que são utilizados para avaliar esses argumentos.” (SKYRMS, 1966, p. 26). A lógica dedutiva utiliza métodos para avaliar a validade argumentativa, e a lógica indutiva métodos para avaliar a força indutiva de argumentos.

Os argumentos dedutivos não se diferenciam dos argumentos indutivos por partirem os primeiros do geral para o específico, enquanto os segundos do específico para o geral. Existem argumentos dedutivos que partem do geral para o geral, do geral para o específico, do particular para o particular e do particular para o geral; os argumentos indutivamente fortes, por sua vez, por exemplo, podem partir do particular para o particular, como no caso da analogia. A diferença principal entre argumentos dedutivamente válidos e indutivamente fortes pode ser encontrada nas definições de força indutiva e validade dedutiva.

Como já visto, a validade dedutiva assegura a correção da conclusão quanto aos termos das premissas – a probabilidade de ser falsa é igual a zero. A probabilidade indutiva, ao seu tempo, “é a probabilidade de que sua conclusão seja verdadeira, uma vez que sejam verdadeiras suas premissas. Assim, a probabilidade indutiva de um argumento é medida da força de evidência que as premissas imprimem à conclusão.” (SKYRMS, 1966, p. 29).

A lógica dedutiva tem conceitos básicos precisos, bem definidos, suas regras foram formuladas rigorosamente. Isso não ocorre com a lógica indutiva – não há regras universais para medir a força indutiva de argumentos. Sabe-se que a lógica indutiva trabalha com probabilidade. A probabilidade indutiva de um argumento é a probabilidade de sua conclusão ser verdadeira, dadas suas premissas.

Uma das maiores funções da lógica indutiva é “esquematizar as expectativas que fazemos relativamente ao futuro, tomando por base o conhecimento do passado e do presente.” Trata-se da questão da previsibilidade. Argumentos indutivos permitem a previsão de situações não cobertas por suas premissas, enquanto que “a lacuna que separa o passado e o presente, do futuro, não pode, portanto, ser coberta pelos argumentos dedutivamente válidos.” (SKYRMS, 1966, p. 36).

A análise de uma grande quantidade de dados por meio das tecnologias de big data é focada principalmente em encontrar correlações – reconhecer padrões, o que remete à lógica indutiva. O pensamento científico tradicional, cartesiano, trabalha principalmente com lógica dedutiva. As formas tradicionais de análise de dados também trabalham com uma espécie de lógica dedutiva de dados. É nesse ponto que o big data diverge tanto das formas tradicionais de análise de dados quanto do pensamento científico clássico, pois funciona a partir de uma lógica indutiva de dados.

Pode-se ilustrar as diferenças de método das lógicas aqui investigadas a partir de um exemplo envolvendo tomada de decisões: imagine a situação na qual uma pessoa decide atravessar a rua. Caso aplique lógica dedutiva, precisará medir todas as variáveis em questão – velocidade de todos os veículos, tempo do sinal de trânsito, dentre outros, e usar cálculos para se chegar no tempo correto de atravessar. Esse processo é lento e dispende muita energia. O cérebro geralmente compreende um contexto global da situação, a partir dos conhecimentos do sujeito, e o processa indutivamente, por meio da generalização probabilística de situações similares. O cérebro integra um número de parâmetros e projeta o possível resultado. Assim, o sujeito não precisa medir todas as variáveis para atravessar a rua, prevê certos comportamentos do trânsito indutivamente e age a partir dessa previsão – uma forma bem mais rápida e econômica de fundamentar a ação racional. Não são necessários parâmetros consistentes e completos, o cérebro se concentra nas informações mais importantes sobre a situação, o que pode produzir erros.

As tecnologias de Big Data são pautadas em uma espécie de lógica indutiva, com a finalidade de guiar o sujeito na complexidade do oceano de dados, possibilitando a busca instantânea por informações cruciais - singularidades na complexidade, correlações ou padrões, o processamento delas como um todo sem precondições, a reprodução efetiva de mecanismos observados no passado e a geração de informações que podem ser utilizadas no presente, guiando ações com mira no futuro. Trata-se de um desafio às próprias bases do conhecimento da ciência clássica cartesiana, fundamentada em uma lógica dedutiva (SCHROFF, 2013).

Não se está afirmando que a indução seja superior, de qualquer modo, à dedução. São duas ferramentas diferentes, com finalidades diferentes. Em alguns casos a dedução é mais eficiente, em outros a indução. A importância está em saber diferenciar a ferramenta mais adequada a cada situação. A dedução, por exemplo, pode ser utilizada para gerar um resultado repetível, enquanto a indução está mais apta a identificar singularidades.

As tecnologias de big data funcionam com algoritmos indutivos, grande parte das vezes a indução nas ciências da computação diz respeito à aplicação do princípio da recorrência a gráficos, um sentido um pouco simplificado da lógica indutiva. Essa lógica indutiva de dados reverte certos parâmetros da lógica dedutiva de dados, frequentemente usados por profissionais:

“‘Store more data for more precision’ becomes ‘forget more data for more possibilities.’ ‘Eliminate singular cases to focus on the most common ones’ becomes ‘abandon frequent cases to focus on differences.’ ‘Model and standardize data’ becomes ‘search for singularities and unknowns.’ ‘Process data exhaustively’ becomes ‘focus on the critical data.’” (MALLE, 2013)⁸

A indução permite rapidez na análise de uma quantidade massiva de dados, com riscos de erro, como visto na própria lógica indutiva. O risco geralmente é aceito em troca da velocidade – uma informação correta e arriscada no momento é preferível a uma completamente confiável tempos depois, principalmente contemporaneamente, onde cada segundo importa no mercado da vida.

Na sociedade de dados ter a informação muitas vezes não é o suficiente – sente-se a necessidade de prever seus impactos sociais. Para essa finalidade, um algoritmo indutivo de velocidade deve ser capaz de antecipar. Para possibilitar isso, a aplicação de computador deve ser implementada em um fluxo contínuo

8 “‘Guardar mais dados para mais precisão’ se transforma em ‘esquecer mais dados para mais possibilidades’. ‘Eliminar casos singulares para se concentrar nos casos mais comuns’ se torna ‘abandonar casos frequentes para se focar na diferença’. ‘Modelar e normalizar dados’ se torna ‘procurar por singularidades e desconhecidos’. ‘Processar dados exaustivamente’ se torna ‘concentrar em dados críticos’.

de dados e reagrupar uma coletividade de agentes responsáveis por transformar a informação o quanto antes em conhecimento utilizável nos instantes seguintes. Desse modo, esse algoritmo deve ser interativo, a informação deve estar disponível rápido e ser refinada em um processo contínuo (AKERKAR, 2014, p. 388).

As tecnologias de big data são uma mudança na maneira como se processa dados. A utilização da indução permite aos algoritmos a reprodução de fenômenos observados, generalizando além das suas fronteiras iniciais. A necessidade de velocidade faz com que os algoritmos tenham que se focar nos dados essenciais. Trata-se de um processo de aprendizagem permanente, nunca completo, para produzir conhecimento imperfeito, porém útil – com a finalidade de prever o futuro e servir como guia no oceano repleto de uma massiva quantidade de dados (SCHROFF, 2013, p. 187-188).

Nesse sentido, as tecnologias do Big Data permitem a navegação no oceano de dados que se está imerso, possibilitando a extração de informações úteis, porém não só isso – põe em debate o próprio modelo científico cartesiano, ao desafiar sua forma arborescente e lógica dedutiva.

3. CONJECTURAS SOBRE A PRÁTICA JURÍDICA NO MUNDO DO BIG DATA.

Observou-se que o mundo passa por transformações intensas. O desenvolvimento acelerado das tecnologias de informação e comunicação nas últimas décadas, como também de automação e controle, desempenharam papel central nessa jornada. Vive-se em uma sociedade onde a informação é um dos bens mais valiosos, possibilitando o exercício de um controle ainda sem medidas exatas ou claras (ainda mais na internet). As tecnologias de big data, nesse cenário, são desenvolvidas com a ambição de possibilitar uma navegação mais rápida e eficiente no oceano de dados produzidos diariamente (e que só tende a crescer) – possibilitando ainda novas modulações de controle.

Nesse espaço de coisas, o mercado e a prática profissional vem se alterando em diversos campos. As bolsas de valores deixaram

de ser lugares onde reinava uma espécie de caos eufórico de compra e venda, e passaram a ser controladas por sistemas informáticos. Os analistas trabalham em suas mesas (principalmente com gráficos), gerindo suas contas a partir do computador. O mercado dos derivativos (no qual a velocidade é essencial) conta cada vez mais com sistemas de inteligência artificial, inclusive realizando operações, com rapidez inalcançável para um ser humano. No mercado de imóveis a internet já é um lugar essencial para a realização de negociações. No de produtos móveis não há o que falar – buscar bens de consumo pela internet virou tarefa corriqueira. Assim, pergunta-se: quais possíveis transformações ocorridas (ou que podem vir a ocorrer) na prática jurídica, tendo em vista a sociedade de controle e informação e as tecnologias de big data?

Parte-se da hipótese de que na atual sociedade de controle e informação o uso das tecnologias de big data gera impactos sobre os mais diversos campos que são tocados pelo fenômeno jurídico, como privacidade, discriminação, direito público à informação, democracia a até automação da tomada de decisão (jurídica), dentre muitos outros. Busca-se no presente construir uma conjectura sobre possíveis transformações na prática jurídica a partir desses impactos.

É importante deixar claro que não se enxerga as novas ferramentas tecnológicas somente como um bem, como uma forma de solucionar os antigos problemas inerentes ao sistema jurídico. Trata-se de uma nova forma de trabalhar, de tratar de problemas, porém não um *Deus ex machina*. Como todo novo método possui vantagens e desvantagens, por vezes até mais desvantagens – diga-se a privacidade quanto ao PRISM revelado por Snowden, desde lá mais ameaçada do que nunca. Nos próximos tópicos serão abordados possíveis transformações na prática jurídica a partir das tecnologias de big data, em ordem de possibilitar conjecturas sobre o assunto.

3.1 BIG DATA COMO FERRAMENTA NA TOMADA DE DECISÃO.

O potencial preditivo das tecnologias de big data, a partir do reconhecimento de padrões em largos conjuntos de dados,

mostra-se como possível ferramenta jurídica na sociedade pautada na informação e no controle. Moses e Chan (2014) afirmam que os métodos analíticos, principalmente nos Estados Unidos, começam a colonizar tanto a administração da justiça quanto a prática jurídica, prometendo respostas sobre a probabilidade de sucesso em uma causa ou sobre a divisão de recursos mais efetiva para o policiamento em uma cidade. Os autores sugerem que “*with sufficiently large datasets and the right analytics and machine learning techniques, we will have simple answers to traditionally difficult questions.*”⁹ (MOSES; CHAN, 2014, p. 643). Assim, os padrões encontrados pela análise podem auxiliar na tomada humana de decisão.

Em seu artigo *Quantitative legal prediction—or—how I learned to stop worrying and start preparing for the data-driven future of the legal services industry* Daniel Martin Katz (2013) afirma que a predição legal quantitativa está cada vez mais próxima da prática jurídica, e irá definir grande parte da inovação na indústria de serviços legais, auxiliada principalmente pelo crescente acesso a imensas quantidades de informações legais semi-estruturadas.

Uma das funções do ator jurídico humano é prever o possível resultado de conflitos legais, geralmente pautando sua estratégia a partir de tal previsão. Treinamento jurídico e experiência são fatores chave para possibilitar essa função. Juristas mais bem treinados e experientes são capazes de prever com mais eficiência o deslinde de casos. Esses profissionais, porém, custam caro e possuem limitações, como todo o ser humano. Katz (2013, p. 928) sustenta que aí entram as tecnologias de previsão, como as tecnologias de big data, que “*are designed to remedy or supplement the shortcomings of human reasoners.*”¹⁰ Assim, não se trata de uma automação completa da decisão, de uma substituição do homem, porém de uma combinação de homem e máquina, com o objetivo de aumentar a eficiência dos processos de tomada de decisão: “*the age of quantitative legal prediction is about a mixture of humans and machines working*

9 “com conjuntos de dados suficientemente grandes e as técnicas certas de análise e aprendizado de máquinas, teremos respostas simples para perguntas tradicionalmente difíceis.”

10 “São criados para remediar ou suplementar as falhas da razão de humanos.”

*together to outperform either working in isolation. The equation is simple: Humans + Machines > Humans or Machines.*¹¹ (KATZ, 2013, p. 929).

Não se está falando de um futuro distante, onde máquinas serão capazes de prever resultados de conflitos humanos – se está falando de tecnologias já existentes. Técnicas de predição já são utilizadas na tomada de decisões, nos mais diversos ramos, públicos e privados. A empresa Lex Machina¹² é um exemplo. Foi criada em 2010 com o objetivo de prever o custo e o resultado de casos jurídicos envolvendo propriedade intelectual, utilizando software com tecnologias de big data. No sistema criminal o estado americano da Virginia utiliza dados estatísticos sobre possibilidade de reincidência criminal para lastrear decisões sobre liberdade condicional, com utilização de sistemas informáticos (MOSES; CHAN, 2014). Algumas jurisdições americanas usam ferramentas de análises de dados para auxiliar em decisões sobre fiança contratual, baseadas em medidas objetivas de risco (MURPHY, 2013). Um modelo bayesiano foi desenvolvido nos Estados Unidos para predição de resultados de acordos em *class actions* envolvendo *securities fraud*. (McSHANE et al, 2012)

A utilização de ferramentas tecnológicas de big data na tomada de decisão certamente foi proporcionada pela evolução do poder computacional (capacidade de coletar e avaliar conjuntos imensos de dados) ocorrida nas últimas décadas. Porém a analítica de dados na área jurídica teve sua origem com o uso de estatísticas e posteriormente de *machine learning* como ferramentas computacionais para tomada de decisão – a chamada “virada empírica”. Trata-se de uma mudança de paradigma na forma da prática jurídica. Decisões legais geralmente são construídas com base no saber jurídico, na experiência e uma certa intuição do aplicador quando a solução

11 “A era da previsão legal quantitativa se trata de uma mistura de humanos e máquinas trabalhando juntos para superar o trabalho de qualquer um dos dois isolados. A equação é simples: Humano + máquina > Humano ou máquina.”

12 “Ele analisa dados ‘training’ e, pelo uso de um algoritmo, identifica a melhor hipótese ligando os dados inseridos às saídas. Os dados ‘training’ são simplesmente exemplos alimentados no algoritmo, dos quais ele aprende potenciais relações preditivas.”

para o caso. Uma visão empírica, como a proposta pela jurimetria de Lee Loevinger (1949), tem como proposta a construção de tais decisões utilizando modelos matemáticos e estatísticos, buscando fundamentá-las em evidências estatísticas. *Machine learning*, por sua vez, como outras técnicas estatísticas, visa inferir certos dados de observações ao longo do tempo. A estatística clássica trabalha com hipóteses construídas a partir de teorias ou pesquisas passadas, testando tais hipóteses com a finalidade de encontrar inferências estatísticas. De modo diverso, *machine learning* não trabalha com hipóteses dadas, “*it analyses ‘training’ data and, through the use of an algorithm, identifies the ‘best’ hypothesis linking input data to outputs. The ‘training’ data is simply the examples fed into the algorithm, from which it ‘learns’ potential predictive relationships.*”¹³ (MOSES; CHAN, 2014, p. 647). Assim, a tecnologia de *machine learning* é dirigida por algoritmos indutivos, buscando a predição estatística de determinados resultados.

Nos anos 70 e 80 as primeiras ferramentas tecnológicas de suporte de decisão jurídica tentavam mimetizar o processo de razão jurídica tradicional. Chamados de *expert systems*, tinham por finalidade a utilização de programas para modelar a forma como um ator jurídico trataria determinado caso, podendo inclusive arazoar suas conclusões (TYREE, 1989). Trata-se de uma tentativa de isomorfismo do ator jurídico – à criação de um espelho do homem. Moses e Chan (2014) argumentam que a maior crítica a esses sistemas se dirige à efetividade, principalmente no que diz respeito a possibilidade de mimetizar corretamente o arrazoamento de um especialista. Os autores afirmam que “*not only are legal rules often contradictory, circular, ambiguous or deliberately vague or contestable, but they rely on social context and human interpretation and cannot be applied directly to raw facts.*”¹⁴ (MOSES; CHAN, 2014, p.

13 “Ele analisa dados ‘training’ e, pelo uso de um algoritmo, identifica a melhor hipótese ligando o dados inserido à saídas. Os dados ‘training’ são simplesmente exemplos alimentados no algoritmo, dos quais ele aprende potenciais relações preditivas.”

14 “Não somente são regras legais muitas vezes contraditórias, circulares, ambíguas ou deliberadamente vagas ou contestáveis, porém elas necessitam de um contexto social e humano de interpretação e não podem ser aplicadas diretamente a dados brutos.”

657). Dessa forma, sistemas que visam auxiliar a tomada de decisão humana seriam mais bem sucedidos na seara jurídica do que aqueles que visam substituir o ator humano. Moses e Chan (2014) atestam que, apesar do insucesso na seara jurídica, os *expert systems* foram utilizados em outras organizações, como na automação de decisões sobre benefícios de bem-estar social nos Estados Unidos – assim, a aceitação de uma tecnologia também depende das características e prioridades de seu grupo alvo. Por outro lado, Moses e Chan (2014) exemplificam os bancos de dados de sentenças e decisões judiciais, algo bastante aceito e difundido no mundo, inclusive no Brasil, como um sistema auxiliar para tomada de decisão humana. Discutem o caso do *New South Wales Judicial Information Research System ('JIRS')*, banco de dados que permite análise estatística de casos criminais passados, auxiliando o juiz na tomada de decisão.

Mais próximo da jurimetria de Loevinger e do *Machine Learning* (se diferenciando deste principalmente pela quantidade de dados analisados), a tecnologia do big data parte de técnicas empíricas, principalmente o método indutivo: “*big data analytics speaks the language of probability, enhancing decision-making by estimating the likelihood that particular facts are or will be true*”¹⁵ (MOSES; CHAN, 2014, p. 663). Desse modo, se diferencia dos *experts systems*, que tentavam mimetizar uma espécie de arrazoamento clássico dedutivo; e dos sistemas de informação e bancos de dados de sentenças, tendo em vista que não buscam dados particulares (mesmo que de forma estatística) em determinado conjunto, e sim padrões e correlações em uma imensa quantidade de dados.

A utilização hoje das tecnologias de big data na prática jurídica é uma continuidade desses primeiros trabalhos. Como já visto, big data é um termo que abrange diversas ferramentas que podem processar imensos volumes de dados, com estruturas diferentes, vindos de fontes diversas. Com a possibilidade de trabalhar cada vez com dados mais complexos, poder-se-ia utilizar as ferramentas para identificar correlações em textos legais, potenciais

15 “*big data analytics* fala a linguagem da probabilidade, melhorando a tomada de decisão pela estimativa da possibilidade de que aquele fato m particular é ou será verdadeiro.”

predições, e até a probabilidade de sucesso em um caso ou a divisão mais igualitárias de um espólio, por exemplo.

Moses e Chan (2014) apresentam um modelo tridimensional para avaliar o uso de ferramentas de big data na seara jurídica: uma dimensão técnica, uma social e outra normativa. A dimensão técnica tem como finalidade considerar questões de funcionalidade e efetividade, ou seja, se as tecnologias de big data são capazes de gerar melhores decisões jurídicas; a dimensão social se presta a observar fatores que podem impactar na disseminação de seu uso; e a terceira, normativa, busca compreender se a tecnologia se adequa à moral dos atores jurídicos e da comunidade em geral. Os autores chamam o primeiro critério de “‘effectiveness’, the second ‘acceptability’ and the third ‘appropriateness’”¹⁶ (MOSES; CHAN, 2014, p. 653-656).

A primeira dimensão, de efetividade, se relaciona à precisão ou exatidão. As informações providas pelas tecnologias de big data devem ser acuradas, em ordem de proporcionar efetividade. Estudos como de Theodor Ruger (2004) mostram tal possibilidade. Nele a predição de um programa de *machine learning*, que utilizava o método de *classification tree*, quanto a votos individuais de *Justices* da Suprema Corte americana em casos futuro foi comparada com predições de professores de direito e advogados bem-sucedidos. O programa de *machine learning* ganhou a disputa de 75% a 59%.

Apesar desse resultado, é importante lembrar que erros estatísticos são comuns. Assim como podem haver acertos, as informações geradas podem ser inúteis para a situação dada. O sistema legal é um sistema complexo adaptativo com níveis elaborados de complexidade e loops extensivos de retromissão entre as respectivas instituições e agentes, assim como instituições e agentes de fora do sistema. O nível de complexidade se diferencia entre os sub-sistemas. Nesse sentido, qualquer predição é somente uma tentativa probabilística, nunca uma certeza (KATZ, 2013).

Além desse fator, Moses e Chan (2014, p. 666) lembram que preconceitos humanos são introduzidos nas análise de dados – “It

16 “‘efetividade’, o segundo ‘aceitabilidade’, e o terceiro ‘adequação’.”

is a human who identifies and selects the data to be analysed and chooses the algorithm to be employed. Some of the time, a human also selects the attributes and variables that are treated as relevant."¹⁷

Não se trata de uma prática guiada por uma racionalidade inteiramente imparcial e totalmente objetiva. Esse é um dos motivos porque que aquele que busca auxílio de tecnologias para tomada de decisão não pode acreditar que os dados, números ou informações geradas falam por si mesmo. São construídos com base em certos preconceitos, e a partir do momento que passam pelo processo de interpretação-compreensão humano ganham contornos de sentido a partir do lócus situacional de cada ator. O sentido não é dado, nem mesmo em um dado – é sempre construído.

Assim, as ferramentas de análise de big data começam a ganhar espaço pelo seu potencial preditivo, porém este, em ordem de ser efetivo, deve ser acurado. Esse potencial de precisão depende de "*data quality, the success of algorithmic approaches (in theory and in practice), the precision with which conclusions are stated and proper testing of conclusions against data independent of the 'training data'*"¹⁸ (MOSES E CHAN , 2014, p. 667). Não se pode acreditar cegamente em benesses trazidas por novas tecnologias – a própria fé na objetividade computacional é quebrada pelos preconceitos do programador. Certos testes foram bem sucedidos, enquanto outros não.

Quanto à dimensão da aceitabilidade, Moses e Chan (2014) consideram difícil prever o nível de aceitação dos profissionais do ramo jurídico, dependendo principalmente dos custos, benefícios e capacidades da tecnologia. O campo jurídico geralmente é relutante quanto a novas tecnologias, se prendendo aos meios tradicionais de resolução de casos. Uma tecnologia que substituísse totalmente o arrazoamento humano, como os *expert systems*, seria de difícil

17 "É um humano que identifica e seleciona os dados a serem analisados e escolhe o algoritmo a ser empregado. Alguns casos, um humano também seleciona os atributos e variáveis que são tratados como relevantes."

18 "qualidade de dados, o sucesso de abordagen de algoritmos (em teoria e prática), a precisão com que conclusão são atestadas e o correto teste de conclusões contra outros conjuntos de dados 'training'."

aceitação em uma visão institucional. Tecnologias, como a analítica de big data, que tem por função auxiliar na decisão humana tem mais chances de aceitação.

A dimensão normativa da adequação gira em torno de problemas como compatibilidade das tecnologias de big data com normas legais, e seu potencial de transparência e *accountability*.

Moses e Chan (2014) afirmam que certos usos de big data podem ser contrários a normas estabelecidas. A razão jurídica clássica leva em consideração fatores normativos para a decisão, enquanto a lógica do big data, indutiva, trabalha com a identificação da correlação de certos fatores com determinados resultados, não precisando ser causalmente relevantes ou ainda normativos. Assim, trata-se de uma porta aberta para decisões discriminatórias, fundadas, por vezes, em fatores ou critérios problemáticos ou historicamente proibitivos, ou até exdrúxulos.

Outras características necessárias para adequação das tecnologias de big data são transparência e *accountability*. Os atores da área jurídica são pouco preparados para lidar com ferramentas tecnológicas, ainda mais com ferramentas de big data. Compreender o funcionamento das complexas técnicas de análise de dados pode ser um problema, gerando dificuldade inclusive de compreensão de resultados. A transparência, assim, é prejudicada. Ademais, com a dificuldade de lidar e traçar o caminho indutivo realizado por tais tecnologias, há riscos quanto ao nível de *accountability* das decisões institucionais (isso foi culpa do computador!).

Há possibilidade de uso das tecnologias de *big data* de forma que auxilie na tomada de decisão, assim como é possível o contrário. É necessário o conhecimento de seu funcionamento e estrutura, em ordem de possibilitar seu uso dentro dos liames propostos, de modo a auxiliar na tomada humana de decisão.

Em uma perspectiva mais geral, parte das funções de todos os profissionais de colarinho branco foram automatizadas, sejam nas finanças, medicina ou direito. Katz (2013) afirma que no futuro de médio prazo a evolução tecnológica ocorrerá no sentido de uma mistura entre máquina e humanos, com a finalidade de realizar trabalhos mais eficientes, que nenhum dos dois poderia efetuar sozinho.

3.2 PROBLEMA DA PRIVACIDADE, BIG DATA E O USUÁRIO-PRODUTO.

O problema entre a privacidade e novas tecnologias não é exclusivo da contemporaneidade. Quando Warren e Brandeis escreveram seu artigo seminal em 1890 – *The Right to Privacy*, a preocupação deles era com as novas tecnologias à época, como máquinas de fotografar e grandes jornais, que, segundo os autores, haviam invadido o sagrado lugar da vida privada doméstica.

Ocorre que na atual sociedade de informação e controle esse problema tem se exacerbado. A tecnologia está cada vez mais acessível e disponível para todos – já são mais de 2.4 bilhões de usuários da internet no mundo – certamente um conjunto muito maior do que Brandeis e Warren imaginavam quando falavam de proteção à privacidade.

Nesse compasso, as legislações nacionais e internacionais ocidentais reiteradamente protegem o direito fundamental à privacidade, como no art. 5º, X, da Constituição Brasileira, porém, por muitas vezes, com enunciados vagos, sem estabelecer critérios claros para proteção, e sobre o que pode ensejar violação a esse direito, principalmente tendo em vista as grandes inovações tecnológicas contemporâneas.

Assim, constrói-se, pelo menos simbolicamente, conceituações para o direito à privacidade. Solove (2013, p. 1880) explica que na atual sociedade de controle e informação “o sistema jurídico provê as pessoas com um conjunto de direitos para possibilitá-las a realização de decisões sobre como gerenciam seus dados.” São direitos primariamente de aviso, acesso e consentimento sobre a coleta, uso e divulgação de dados. Tais direitos têm como finalidade possibilitar o controle dos indivíduos sobre suas informações pessoais. Dessa forma, por meio desse controle, ter liberdade para decidir como sopesar os benefícios e os custos da coleta, uso e divulgação de suas informações – Solove chama essa concepção de “privacy self-management”, a qual o presente trabalho se referirá como autogerenciamento da privacidade.

O conceito de autogerenciamento da privacidade, de acordo com Solove (2013), é albergado no consentimento do sujeito,

buscando neutralidade sobre a substância, ou seja, se aquela forma específica de coleta, uso ou divulgação é boa ou ruim, e mirando sobre o consentimento do sujeito em inúmeras práticas de privacidade.

Parece que a concepção do direito à privacidade como autogerenciamento da privacidade, fulcrado no consentimento individual, vem se ajustando como a forma presente de se compreender tal manifestação social no contexto da sociedade de informação e controle.

Faz-se ainda uma diferenciação entre informações que gozariam de mais proteção e informações que necessitam de menos tutela, separando-as como dados pessoais (não sensíveis) e dados sensíveis. Tal diferenciação serve como critério para aferir violações à privacidade – a utilização de um dado sensível sem o consentimento do titular constitui uma ofensa a sua privacidade, porém a mesma situação quanto a um dado pessoal pode não ser considerada uma violação, dependendo do contexto (DONEDA, 2006).

A diferenciação em questão se trata de mais uma forma de tentar criar critérios para aferir a violação de privacidade em um caso concreto, tendo em vista que dados sensíveis merecem proteção maior que dados pessoais, que servem meramente para a identificação do indivíduo. Porém estes últimos também gozam de proteção, não podendo ser utilizados de forma arbitrária ou lesiva.

Nesse sentido, da privacidade como autogerenciamento informacional, a coleta, utilização e divulgação de dados sem o consentimento do sujeito interessado consiste em uma violação ao seu direito à privacidade, logo uma invasão à sua vida privada. Deve-se esclarecer que este é um conceito histórico, construído para, de certa forma, manter a autonomia do indivíduo sobre sua própria vida dentro de uma sociedade onde a informação é um bem extremamente valorizado.

Escândalos sobre violação de privacidade não faltaram nos últimos anos. Em todos os âmbitos. Públicos e privados. Familiares e sociais. Nunca se foi tão vigiado. O caso PRISM talvez retrate essa situação na qual se vive. No artigo *big data in small hands*, Hartzog e Selinger (2013) argumentam que os benefícios sociais das tecnologias de big data devem ser conciliados com o risco que

cada vez mais apresentam para a privacidade individual. Quanto mais informações são coletadas sobre indivíduos, hoje quanto à localização, finanças, uso de eletricidade, saúde, atividade online, dentro outros, “*concerns arise regarding profiling, tracking, discrimination, exclusion, government surveillance, and loss of control.*”¹⁹ (HARTZOG, SELINGER, 2013, p. 251).

Desconfia-se da capacidade da construção da privacidade como autogerenciamento possibilitar um controle significativo da vida privada na sociedade da informação. Talvez o caminho seja uma desconstrução do próprio entendimento da privacidade como autogerenciamento, principalmente no que tange a capacidade de consentimento racional, possibilitando a construção de um conceito mais adequado à proteção de direitos individuais.

Em seu artigo *Abandon all Hope?* Graham Greenleaf (2014, p. 640) trata sobre alguns fatores que dificultam a preservação da privacidade contemporaneamente: fatores ambientais, “*these include the progressive elimination of processing costs (Moore’s law); the progressive elimination of data storage costs; and the explosion of online transactions*”; a imensa geração e coleta de dados pessoais, o que envolve “*digitisation of new/more personal data; visual data collection; voluntary disclosure; big data aggregations; de-anonymisation of transactions; re-identification techniques; government IDs; and commercial personal data collection.*”; a intensificação do uso e processamento de dados pessoais, por meio de diversas ferramentas construídas com esse intuito “*including by more powerful processing of personal data (analytics, search and ranking, etc); through commercial interconnection (eg, advertising syndication, internet of things); and by state interconnection.*”; a redução da segurança dos dados pessoais e aumento de sua retenção, “*including by permanent retention of personal data (both on our own devices, and on third-party devices); by the endemic failure of security systems; and the risks of massive systems failure and data unavailability.*”; e a crescente transferência e divulgação de dados pessoais, que ocorrem geralmente por “*state surveillance*

19 “preocupações começam a surgir sobre *profiling, tracking*, discriminação, exclusão, vigilância do governo e perda de controle.”

(Snowden revelations); foreign state surveillance (Snowden again); international data mobility (cloud services, etc); malicious hacking (markets in stolen personal data, botnets); and unintended data breaches.”

Existem riscos inerentes à utilização das tecnologias preditivas, de acordo com Hartzog e Selinger (2013, p. 252), *“the relegation of decisions about an individual’s life to automated processes based on algorithms and artificial intelligence raises concerns about discrimination, selfdetermination, and the narrowing of choice.”*

Nesse sentido, em um mundo de controle e informação, respaldado por massivos sistemas de dados, torna-se cada vez mais difícil resguardar a privacidade, quiçá o próprio indivíduo, que se tornou produto no comércio da internet, que segue a máxima: quando o serviço é grátis, o usuário é o produto. Como lembra Greenleaf (2014, p. 641), *“the aggregation of information about individuals for the purpose of onselling that information for other marketing activities has come to predominate.”*²⁰

Hartzog e Selinger (2013) reafirmam essa preocupação de outro modo: as tecnologias de big data beneficiam organizações e não indivíduos, pois esses são o produto – afinal, se não se está pagando o serviço, não se é consumidor, porém o produto. Em interações online, informações pessoais são trocadas por serviços gratuitos. A partir daí, organizações passam a conhecer todas as preferências dos indivíduos, oferecendo produtos que se adequem a elas, dentre outras possibilidades. Nesse sentido, as organizações não estão preparadas para dividir a riqueza criada pelas informações pessoais com os indivíduos.

Crisan, Zbucnea e Moraru (2014) afirmam que a segurança de dados pessoais online não é mais possível. Todas as atividades de um indivíduo online são submetidas à monetização – no modelo em que ele não é mais um consumidor, e sim o produto. Os autores defendem que tanto as empresas quanto os próprios indivíduos deveriam ser mais responsáveis com informações pessoais e sensíveis,

20 “A agregação de informações sobre indivíduos com o objetivo de venda dessas informações para outros ramos do mercado domina as práticas.”

tendo em vista a impossibilidade de segurança online.

Por outro lado, os sistemas de direito à privacidade continuam a ser reforçados. Na Europa, por exemplo, o Regulamento 12 e a Convenção 108 tem a função de combater as ameaças à privacidade, incluindo aquelas de fora do continente, por meio de “*more explicit data minimisation, the ‘right to be forgotten’, data portability, ‘privacy by default’, stronger extraterritoriality, local representative requirements, and fines proportional to business size.*”²¹(GREENLEAF, 2014, p. 642). Tentativas como essas, segundo Greenleaf (2014), são rotineiramente sem sucesso, em razão da habilidade das empresas americanas, juridicamente fundada ou não, de adquirir, processar e usar dados de pessoas de todo o mundo com poucas restrições. “*Privacy standards in other countries do not matter much if personal data can be liberated to the US ‘Safe Harbor’.*”²² (GREENLEAF, 2014, p. 642).

Vive-se em uma tensão constante, territorialização e desterritorialização, controle e liberdade. O modelo predominante na economia de informação hoje é do usuário-produto. Deve-se recordar que existem nichos e locais onde esse modelo não se aplica, sejam atuais ou atualizáveis (virtuais), mas isso não impede que seja o modelo dominante – pelo contrário, por muitas vezes a exceção somente confirma a regra. Greenleaf (2014, p. 641-642) adverte que tal modelo predominante não é necessariamente permanente, mas para não o ser depende ou do estouro de uma segunda bolha da internet, ou “*a concerted effort by the rest of the world to reject privacy invasive business practices. Neither is impossible, nor likely to occur rapidly.*”

21 “minimização de dados mais explícita, o direito ao esquecimento, portabilidade de dados, privacidade por padrão, extraterritorialidade mais forte, requerimento de representates locais e multas proporcionais ao tamanho do negócio.”

22 “Padrões de privacidade em outros países não importam tanto quando dados pessoais podem ser liberados para o porto seguro dos EUA.”

3.3 OS PARADOXOS E A ÉTICA DE BIG DATA.

No artigo *Three paradoxes of big data* Richards e King (2013) tratam sobre três pontos aporéticos sobre benefícios e malefícios das tecnologias de big data. O primeiro paradoxo fala sobre transparência. As tecnologias de big data prometem usar dados coletados para conhecer mais o mundo, deixá-lo mais transparente, porém, ao mesmo tempo, promovem coletas invisíveis, utilizando ferramentas opacas, rodeadas de mistério. Os autores perguntam: “*If big data spells the end of privacy, then why is the big data revolution occurring mostly in secret?*”²³ (RICHARDS; KING, 2013, p. 42).

O segundo é o paradoxo da identidade. Richards e King (2013) afirmam que ao mesmo tempo que tenta identificar, o big data também ameaça a identidade. Na maioria das democracias liberais tem-se o direito de escolha sobre identidade – cada pessoa pode escolher o que deseja ser (pelo menos pretensamente sob o véu do Estado de Direito). Com as tecnologias de big data, e a combinação de dados telefônicos, histórico da internet, histórico de compras, posts em redes sociais, dentre outros, é possível que a escolha de identidade se torne uma imposição. O “eu sou” se torna “você é”, e o “eu gosto” se torna “você gosta”.

Por último, o paradoxo do poder. Diz-se que as tecnologias de big data permitem uma visão diferente do mundo, que pode a todos beneficiar. Nesse novo modelo de mundo existirão aqueles mais beneficiados e aqueles menos beneficiados (ou até maleficiados). Richards e King (2013) afirmam que certamente aqueles mais beneficiados serão instituições que controlam ferramentas de big data sobre indivíduos, ou indivíduos, que tem seus dados constantemente minerados, analisados e agrupados. A penumbra da tecnologia, sem claras limitações legais e opacas limitações técnicas, faz com que tanto os indivíduos quanto as instituições se movam em terreno movediço. Muitos indivíduos continuam sob a

23 Se big data traz o fim da privacidade, porque a revolução de big data está acontecendo em sua maior parte em segredo?”

penumbra do discurso institucional da tecnologia como *Deus ex machina*, dos serviços gratuitos por pura bondade e segurança do ambiente online. As organizações continuam a utilizar informações de maneira excusa, legalóide, sob o risco de escândalos, como o PRISM. O resultado dessa relação é “*an uneasy, uncertain state of affairs that is not healthy for anyone and leaves individual rights eroded and our democracy diminished.*”²⁴ (RICHARDS; KING, 2013, p. 45).

Nesse sentido, Richards e King (2014) afirmam que as tecnologias de big data vem aumentando o poder de organizações, principalmente no Ocidente, o que requer o desenvolvimento de uma espécie de ética do big data. Argumentam que conceitos humanos como identidade, privacidade, liberdade de escolha, transparência e confidencialidade devem ser balanceados com o uso do big data, sob o risco de cessarem de existir. Essa é a missão de uma ética de big data.

Richards e King (2014) desenvolvem a ética do big data como quatro princípios que deveriam governar o fluxo de dados na sociedade de controle e informação. Em primeiro lugar deve-se reconhecer a privacidade como regra do fluxo de informações. Deve-se reconhecer, em segundo lugar, que informações pessoais sensíveis compartilhadas podem permanecer privadas. Em terceiro lugar deve-se reconhecer a necessidade de transparência nas práticas de big data. Por último, os autores afirmam a necessidade do reconhecimento de que big data pode comprometer a identidade de indivíduos “*by allowing institutional surveillance to identify, categorize, modulate, and even determine who we are before we make up our own minds.*”²⁵ (RICHARDS; KING, 2014, p. 396).

24 “Um incerto estado de coisas que não é saudável para ninguém e deixa direitos dos indivíduos em erosão e nossa democracia diminuída.”

25 “Ao permitir a vigilância institucional identificar, categorizar, modular e até determinar quem nós somos antes de nós próprios escolhermos.”

4. CONSIDERAÇÕES FINAIS.

Contemporaneamente as tecnologias da comunicação e informação (TIC) vêm modificando as relações humanas. Seja pela velocidade na comunicação e no acesso a informação, pelos os novos hábitos consequentes das redes sociais ou pelo constante controle de informação. Nessa breve conjectura objetivou-se compreender como a tecnologias de *big data* podem modificar (ou já modificam) a prática jurídica, dentro da presente sociedade de controle e informação. Teve por hipótese que o uso das tecnologias de big data gera impactos sobre os mais diversos campos que são tocados pelo fenômeno jurídico, como privacidade, discriminação, direito público à informação e até automação da tomada de decisão (jurídica), dentre muitos outros. Buscou-se construir uma conjectura sobre possíveis transformações na prática jurídica a partir desses impactos.

Apresentou-se as tecnologias de big data. Com a rápida evolução das tecnologias da informação e comunicação (TIC) nos últimos dez anos, a possibilidade de criar conteúdo foi incrivelmente amplificada. É estimado que entre a invenção da escrita e 2006 a humanidade acumulou cerca de 180 Exabytes de dados. Esse número cresceu dez vezes entre 2006 e 2011, chegando a um total de 1600 Exabytes. Perguntou-se de que modo poderia se navegar de modo mais eficiente nesse oceano de dados recriado diariamente. Foi visto que a tecnologia emergente do big data foi criada com o intuito de buscar por correlações nesse oceano de dados – outra maneira de se guiar nesse espaço de eterna criação e recriação, com a possibilidade de processar rapidamente grandes quantidades de dados, por vezes heterogêneos, com a finalidade de extrair informações úteis. A lógica indutiva, por sua vez, a ferramenta utilizada para dar vazão a eficiência e rapidez, em contraposição a clássica lógica dedutiva.

Buscou-se compreender como as ferramentas de big data e analytics podem transformar a prática jurídica. Partiu-se da hipótese de que na atual sociedade de controle e informação o uso das tecnologias de big data gera impactos sobre os mais diversos campos que são tocados pelo fenômeno jurídico, como privacidade,

discriminação, direito público à informação e até automação da tomada de decisão (jurídica), dentre muitos outros. Buscou-se construir uma conjectura sobre possíveis transformações na prática jurídica a partir desses impactos. Foi visto que há possibilidade de uso das tecnologias de *big data* de forma que auxilie na tomada de decisão, assim como é possível o contrário. É necessário o conhecimento de seu funcionamento e estrutura, em ordem de possibilitar seu uso dentro dos liames propostos, de modo a auxiliar na tomada humana de decisão. Por outro lado, foi observado que as tecnologias de big data podem criar novos tipos de discriminação, além de ameaçarem cada vez mais a privacidade e até o processo pessoal de construção de identidade. Foram vistos certos paradoxos na tecnologia, que podem sofrer menos tensão a partir da construção de uma ética do big data. Como já avisava aquele antigo grupo de rock inglês, *welcome to the machine*.

REFERÊNCIAS

- AKERKAR, Rajendra. *Big Data Computing*. London: CRC, 2014.
- BOYD, D. and K. Crawford. *Critical questions for big data*. Information, Communication & Society 15(5): 662–679, 2012.
- CARNAP, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1963.
- CRIȘAN, C., ZBUCHEA, A., MORARU S. *Big Data: The Beauty or the Beast*. Strategica: Management, Finance, and Ethics, Bucareste, n. 2, 2014.
- DEAN, J.; S. GHEMAWAT. *MapReduce: Simplified data processing on large clusters*. Communications of the ACM 51(1): 107–113, 2008.
- DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.
- ERMOLAYEV V., AKERKAR R., TERZIYAN V, COCHEZ M. *Toward Evolving Knowledge Ecosystems for Big Data Understanding*. In: AKERKAR, Rajendra (Ed.). *Big Data Computing*. Sognal, Norway: CRC, 2014.

FAN, W.; BIFET, A.; YANG, Q.; YU, P. *Foreword*. In *Proc First Int Workshop on Big Data, Streams, and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*. New York: ACM., 2012.

FISHER, D.; DELINE R.; CZERWINSKI M.; DRUCKER S. *Interactions with big data analytics*. *Interactions* 19(3):50–59, 2012.

FLORIDI, Luciano. *Big data and Their Epistemological Challenge*. In: *Philosophy and Technology*. Vol. 25, issue 4. Springer Science, 2012.

GRELLER, W. *Reflections on the knowledge society*. Disponível em <<http://wgreller.wordpress.com/2010/11/03/big-data-isnt-big-knowledge-its-big-business/>, 2012>. Acesso em 18/01/2014.

GREENLEAF, Graham. *Abandon all Hope?* UNSW Law Journal, Sydney, Australia, v. 37, n.2, p.636-642, 2014.

HARTZOG, W.; SELINGER, E. *Big data in small hands*. *Stanford Law Review*, Califórnia, v. 66, n. 2, 2013.

KATZ, D. M. *Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry*. *Emory Law Journal*, Atlanta, v. 62, n. 2, 2013.

LIN, J.; DYER, C. *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Synthesis Lectures on Human Language Technologies, 2012.

LOEVINGER, Lee. *Jurimetrics: The Next Step Forward*, *Minnesota Law Review*, Minnesota, v. 455, n. 33, 1949.

MALLE, Jean-Pierre. *Big Data: farewell to Cartesian thinking?* Disponível em: <<http://www.paristechreview.com/2013/03/15/big-data-cartesian-thinking/>>. 2013.

MANYIKA, J.; CHUI, M.; BROWN, B.; et al. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011.

McSHANE B. B.; WATSON. *Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits*. *Journal of Empirical Legal Studies*, Pennsylvania, v. 9, n.1, p. 482-505, 2012.

MOSES, L. B.; CHAN, J. *Using Big Data For Legal And Law Enforcement Decisions: Testing The New Tools*. UNSW Law Journal, Sydney, Australia, v. 37, n.2, p.643-678, 2014.

MURPHY, Tonia Hap. *Mandating Use of Predictive Coding in Electronic Discovery: An Ill-Advised Judicial Intrusion*. American Business Law Journal, Chicago, v. 50, n. 3, p. 609-657, 2013.

PUGLIESI, Marcio. *Teoria do Direito*. 2 ed. São Paulo: Saraiva, 2009.

REICHENBACH, Hans. *Erfahrung und Prognose*. Springer Fachmedien Wiesbaden, 1983.

RICHARDS, N. M.; JONATHAN H. *Three Paradoxes of Big Data*. Stanford Law Review, Califórnia, v. 41, n. 2, 2013.

_____, *Big Data Ethics*. Wake Forest Law Review, Winston-Salem, v. 89, n. 4, 2014.

RUGER, T. W.; KIM, P. T.; MARTIN, D. A.; QUINN, K. M. *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*. Columbia Law Review, New York, v. 1150, n. 104, 2004.

SCHROFF, Gautam. *The intelligent web: search, smart algorithms and big data*. Oxford: Oxford university press, 2013.

SKYRMS, Bryan. *Escolha e acaso*. São Paulo: Cultrix, 1966.

SOLOVE, Daniel J. *Privacy Self-Management and the consent dilemma*. Harvard Law Review, vol. 126. MA: Harvard University Press, 2013.

TYREE, Alan. *Expert Systems in Law*. Australia: Prentice Hall, 1989.

WARREN, Samuel; BRANDEIS, Louis. *The right to privacy*. Cambridge, MA: Harvard Law Review, Vol IV, N. 5, 1890.

Recebido em 24/04/2015.

Aprovado em 02/06/2015.