

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

VINICIUS LANDIN MACEDO

**O USO DO *BIG DATA* NO CONTEXTO DA ECONOMIA DIGITAL SOB A ÓTICA
DAS FALHAS DE MERCADO**

**Porto Alegre
2021**

VINICIUS LANDIN MACEDO

**O USO DO *BIG DATA* NO CONTEXTO DA ECONOMIA DIGITAL SOB A ÓTICA
DAS FALHAS DE MERCADO**

Trabalho de conclusão de curso de graduação apresentado ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Ciências Econômicas.

Orientador(a): Prof. Dr. Stefano Florissi

Porto Alegre

2021

CIP - Catalogação na Publicação

Macedo, Vinicius Landin
O uso do Big Data no contexto da economia digital
sob a ótica das falhas de mercado / Vinicius Landin
Macedo. -- 2021.
98 f.
Orientador: Stefano Florissi.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Ciências Econômicas, Curso de Ciências Econômicas,
Porto Alegre, BR-RS, 2021.

1. Falhas de mercado. 2. Big data. 3. Economia
digital. 4. Competição. I. Florissi, Stefano, orient.
II. Título.

VINICIUS LANDIN MACEDO

**O USO DO *BIG DATA* NO CONTEXTO DA ECONOMIA DIGITAL SOB A ÓTICA
DAS FALHAS DE MERCADO**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para a obtenção do título de Bacharel em Ciências Econômicas.

Aprovado em: Porto Alegre, __ de _____ de 2021

BANCA EXAMINADORA:

Prof. Dr. Stefano Florissi - Orientador

UFRGS

Prof. Dr. Maurício Andrade Weiss

UFRGS

Prof. Dr. Leonel Toshio Clemente

UFRGS

AGRADECIMENTOS

Primeiramente, gostaria de agradecer meus pais por terem batalhado tanto para fornecer um ambiente de constante aprendizado, dentro e fora de casa, e dar as melhores condições e oportunidades para o meu desenvolvimento pessoal e acadêmico. Espero que todos os anos que passei longe do lugar onde cresci e de minha família façam jus à nobre motivação.

Gostaria, também, de agradecer a todos os colegas e amigos que conheci na universidade, compartilhando esta jornada comigo, e que agora tenho a honra de finalizar o ciclo conjuntamente. Citação especial para Carlos, Clara, Gabriel, Pedro e Vitor. À minha companheira, Priscila, que me acompanhou desde o início da escrita do trabalho, sempre apoiando e acreditando em mim, gerando um impacto extremamente positivo e incentivador na busca do melhor de minha capacidade. Agradeço imensamente toda sua paciência e amor dedicados a mim, antes e durante a escrita, pois me ajudaram a ser o que sou hoje e chegar até aqui. Ao meu cachorro, Milow, por ter passado toda a escrita deste trabalho literalmente ao meu lado, passando por uma exaustiva quarentena e longas madrugadas de ansiedade juntos.

Agradeço a todos os professores que me fizeram lembrar o porquê de eu ter ingressado em uma graduação de Economia, ministrando aulas tão agregadoras e que com certeza irei levar para a vida toda, seja como meras reflexões acadêmicas ou como inspiração pessoal. À banca examinadora, professores Maurício Weiss e Leonel Toshio, por terem se interessado pelo trabalho e aceitarem dar sua contribuição. Por fim, mas não menos importante, agradeço imensamente ao meu orientador, professor Stefano Florissi, por ter me acolhido tão bem, acreditar em minha capacidade, dar ótimos conselhos que lapidaram o trabalho final e por ser tão querido sempre.

RESUMO

A rápida emergência do fenômeno do *big data* e a digitalização dos produtos e serviços ao redor do mundo trouxeram consigo uma mudança na dinâmica de interação entre os agentes econômicos e na forma em que a informação é aplicada na economia. O presente trabalho, portanto, objetiva compreender como as falhas de mercado atuam neste novo contexto da economia digital, apresentando os possíveis desequilíbrios distributivos na interação entre firma e consumidor. Para isto, utiliza-se uma análise exploratória com revisão de literatura teórica e empírica, junto a técnicas de pesquisa bibliográfica e documental. Depreende-se do trabalho que existe uma grande sustentação na teoria econômica sobre a atuação de quatro principais fenômenos sobre a economia digital: assimetria de informação, externalidades, poder de mercado e discriminação de preços. Embora exista uma carência de acompanhamento da teoria por estudos empíricos que evidenciem estas falhas de fato, autoridades governamentais, incluindo o governo do Brasil, reconhecem-nas como perigos eminentes e buscam soluções via regulamentações de dados. O contexto atual demonstra um grande potencial para exploração de novas evidências e experimentações sobre os efeitos das embrionárias regulamentações sobre os dados e como os indícios de falhas de mercado são afetados por elas.

Palavras-chave: Falhas de mercado. *Big data*. Economia digital. Competição.

ABSTRACT

The fast emergence of the big data phenomenon and the products and services digitization around the world happened with changes in the dynamic of interaction between economic agents and in the way in which information is used in the economy as an input. The present work aims to understand how market failures act in the new context of the digital economy, presenting possible distributive disequilibrium in the interaction between the firm and the consumer. For this purpose, an exploratory analysis is used with a review of theoretical and empirical literature, along with bibliographic and documentary research techniques. It's concluded from the work that there is great support in economic theory about the action of four main phenomena on the digital economy: information asymmetry; externalities; market power; and price discrimination. Although there is a lack of theory supported by empirical studies that evidence these failures in fact, government authorities, including Brazil, already recognize them as imminent dangers and seek for solutions through data regulations. The current context shows great potential to explore new evidence and experimentation about the effects of new regulations on data and how the signs of market failures is affected by them.

Keywords: Market failures. Big data. Digital economy. Competition.

LISTA DE FIGURAS

Figura 1 – Custo médio de armazenamento de dados por consumidor (1998-2012) em dólares por gigabyte	15
Figura 2 – Proporção da população utilizando a Internet* (1990-2018), comparativo mundo e Brasil.....	16
Figura 3 – Capacidade mundial de armazenamento de informação em bytes (esquerda) e crescimento da capacidade de telecomunicação em kbps (direita).....	18
Figura 4 – Capitalização de mercado das maiores empresas <i>data-driven</i> (2019) em bilhões de dólares	21
Figura 5 – Média de receita por empregado entre as 250 maiores empresas de TIC (2000-2013) em um mil de dólar	22
Figura 6 — Resposta da pesquisa “ <i>The Business of Data</i> ”	24
Figura 7 – Pirâmide do conhecimento	30
Figura 8 – Definições de “ <i>big data</i> ”	34
Figura 9 – O uso de dados: modelos de negócio <i>data-enabled</i> vs <i>data-enhanced</i>	38
Figura 10 – Dados de <i>input</i> e <i>output</i> para plataformas digitais	39
Figura 11 – <i>Data Value Chain</i> (Cadeia de valor dos dados)	41
Figura 12 – Da produção linear ao <i>feedback loop</i> na economia digital.....	42
Figura 13 – <i>Big data value chain</i> (Cadeia de valor do <i>big data</i>).....	43
Figura 14 – Fatores que influenciam na determinação do valor dos dados.....	45
Figura 15 – Características das estruturas de mercado dos dados	46
Figura 16 – Valorando a correlação dos dados entre dois usuários.....	55
Figura 17 – Ilustração da precificação personalizada.....	58
Figura 18 – Posições dominantes e <i>market share</i> das GAFM (2019-2020) em %	68
Figura 19 – Indicadores combinados de <i>market share</i> das duas atuais empresas líderes em determinados mercados digitais no Reino Unido (2010 - 2018)	69
Figura 20 – Aumento do poder de mercado no setor de tecnologia (mudança percentual acumulativa).....	70
Figura 21 – Diferenças de preços da Amazon baseado na localização geográfica do consumidor, utilizando os preços de Nova Iorque (EUA) como referência.....	72
Figura 22 – Simulação das mudanças dos resultados devido à personalização de preços	73
Figura 23 – Simulação das mudanças dos resultados devido à personalização de descontos (discriminação de preços de segundo grau otimizada)	73

Figura 24 – Triângulo do excedente	74
Figura 25 – Usuários de <i>ad-block</i> em <i>mobile</i>	77

SUMÁRIO

1 INTRODUÇÃO	11
2 CONTEXTUALIZAÇÃO E ESTRUTURAÇÃO DE UMA ECONOMIA GLOBAL DIGITALIZADA BASEADA EM DADOS DIGITAIS	13
2.1 UM BREVE HISTÓRICO DO DESENVOLVIMENTO TECNOLÓGICO DIGITAL	13
2.2 A EVOLUÇÃO DA INFORMAÇÃO DIGITAL ATÉ O <i>BIG DATA</i> E A FORMAÇÃO DA ECONOMIA <i>DATA-DRIVEN</i>	16
2.3. APLICAÇÕES DOS DADOS DIGITAIS NO SÉCULO XXI	23
2.4 NOVAS IMPLICAÇÕES ÉTICAS NO CONTEXTO DIGITAL.....	25
3 DEFINIÇÃO DE ESCOPO CONCEITUAL	28
3.1 CARACTERIZAÇÃO DOS DADOS DIGITAIS	28
3.1.1 Definição	28
3.1.2 Classificação	32
3.2 CARACTERIZAÇÃO DO <i>BIG DATA</i>	33
3.3 CARACTERIZAÇÃO DA ECONOMIA <i>DATA-DRIVEN</i>	35
3.3.1 Modelos de negócio: <i>data-enhanced</i> vs <i>danta-enabled</i>	37
3.3.2 Plataformas digitais	38
3.3.3 Cadeia de valor dos dados (“<i>Data value chain</i>” e “<i>Big data value chain</i>”)	40
3.4 PRECIFICAÇÃO DOS DADOS DIGITAIS.....	43
3.4.1 Estrutura de mercado e estratégias de preço	45
4 AS FALHAS DE MERCADO NO USO DO <i>BIG DATA</i> DENTRO DA ECONOMIA DIGITAL	48
4.1 EMBASAMENTO TEÓRICO	48
4.1.1 Características econômicas dos dados	48
4.1.2 Assimetria de informação	51
4.1.3 Externalidades	52

4.1.4 Poder de mercado	56
4.1.5 Discriminação de preços.....	58
4.2 REVISÃO EMPÍRICA	59
4.2.1 Assimetria de informações: caso das ferramentas de pesquisa <i>online</i> e do modelo de consentimento informado	60
4.2.2 Externalidades: caso dos algoritmos de <i>machine learning</i> e os mecanismos de proteção à privacidade online.....	62
4.2.3 Poder de mercado: caso das GAFAMs	67
4.2.4 Discriminação de preços: caso da Amazon, Netflix e ZipRecruiter	71
4.3 COMO OS ATORES ECONÔMICOS LIDAM COM AS FALHAS DE MERCADO	75
4.3.1 Consumidores.....	75
4.3.2 Governo.....	78
5 CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS.....	85

1 INTRODUÇÃO

A transformação da economia e de suas estruturas de funcionamento teve um grande impacto pelo surgimento do fenômeno do *big data*, que pautou a importância econômica dada aos dados digitais como forma de agregar valor a modelos de negócios. Esses dados valiosos, no entanto, são baseados em informações pessoais de consumidores, coletados por diversos mecanismos ao longo da navegação da internet e transmitida por uma grande rede de comunicação. A interação entre empresas digitais e usuários pode gerar efeitos distributivos adversos, criando estruturas de falhas de mercado específicas do contexto da economia digital.

O presente trabalho tem como objetivo observar de que forma os fenômenos econômicos, mais especificamente as falhas de mercado, atuam sob o contexto da economia digital e do *big data*, analisando como se dá a interação entre os agentes econômicos e quais as evidências empíricas que a sustentam. É de grande importância entender de que forma esta dinâmica ocorre e como pode ser utilizada como ferramenta geradora de desequilíbrios e injustiças entre os consumidores – ilustrado por exemplos da realidade recente como os que ocorreram no escândalo *Cambridge Analytica* e o vazamento de dados da empresa Netshoes no Brasil –, a fim de auxiliar a criação de mecanismos regulatórios para mitigar tais efeitos.

Além disso, ainda existe um grande obscurantismo sobre a atuação dos mercados digitais e em relação à forma que as firmas lidam com os dados dos usuários e consumidores. Em recente levantamento produzido pela KPMG, “Global Customer Experience Excellence Research 2020”, 98% dos entrevistados disseram que estão preocupados com seus dados pessoais e a qual destino eles são utilizados. Portanto, é necessário analisar a possível existência de assimetrias de informação entre os agentes econômicos, de que forma atuam no equilíbrio distributivo e quais são seus efeitos sobre outras falhas de mercado e aos consumidores.

A metodologia a ser utilizada é a análise exploratória com revisão de literatura teórica e empírica sobre os fenômenos estudados, dentro do contexto da economia digital. Serão aplicados alguns conceitos da economia da informação e da microeconomia, observados por meio de técnicas de pesquisa bibliográfica e documental.

O trabalho será dividido em três capítulos e objetivos específicos, além da introdução e das considerações finais. O segundo capítulo buscará construir um panorama histórico sobre a digitalização da economia, o surgimento do *big data* e o que mudou na dinâmica econômica. O terceiro conceituará os fenômenos integrantes da economia digital, a fim de limitar o escopo conceitual e entender de que forma são aplicados no contexto da economia. O quarto capítulo colocará em contexto os fenômenos da economia digital sob a ótica das falhas de mercado,

como são gerados atritos distributivos e de que forma os atores econômicos lidam com este cenário. Para isso, será realizado um levantamento da literatura empírica a fim de avaliar se o embasamento teórico é robustamente sustentado por evidências. Adicionalmente, será observado o atual cenário em relação às ações dos outros agentes econômicos além da perspectiva da firma e de forma eles lidam com os fenômenos econômicos a serem explicitados anteriormente no capítulo.

2 CONTEXTUALIZAÇÃO E ESTRUTURAÇÃO DE UMA ECONOMIA GLOBAL DIGITALIZADA BASEADA EM DADOS DIGITAIS

Para entender de que forma ocorreu a transformação da dinâmica econômica, chegando ao atual panorama praticamente estabelecido sobre a digitalização dos produtos e serviços e do uso dos dados digitais dentro da economia, é importante também compreender como se deu este processo evolutivo, colocando o objeto de estudo em seu devido contexto histórico. Neste capítulo serão observadas as principais questões que explicam sucintamente de que forma os instrumentos tecnológicos que sustentam a economia digital se expandiram dentro da sociedade, como a informação chegou ao patamar de bem de troca e as diferenças na aplicação de dados antes do fenômeno do *big data*, além das novas implicações éticas resultantes desta dinâmica.

2.1 UM BREVE HISTÓRICO DO DESENVOLVIMENTO TECNOLÓGICO DIGITAL

O rápido processo de desenvolvimento dos meios de comunicação e da eficiência na troca de informações nas últimas décadas é resultante, antes de tudo, da história da computação desde meados da década de 1940, durante o desenvolvimento e comercialização de mecanismos tecnológicos para uso militar na Segunda Guerra Mundial (CERUZZI, 2003). Neste contexto, o enfoque está nas máquinas que realizam cálculos de forma extremamente rápida, gerando uma grande quantidade de informações em pouco tempo. O complemento técnico do advento destas máquinas vem na década seguinte com as novas tecnologias de armazenamento, que podiam processar a informação digitalizada de forma mais abrangente. A partir dessas duas grandes inovações observa-se o primeiro fenômeno econômico sobre a informação vista como um bem digital: a diminuição do seu custo marginal de reprodução (GOLDFARB; TUCKER, 2019).

Com este ambiente, no qual a informação é passível de ser digitalizada e reproduzida a baixíssimo custo econômico, desenvolvem-se as primeiras gerações de computadores comerciais, ainda vendidos em quantidades modestas nos anos 1960 (CERUZZI, 2003). Entretanto, é importante ressaltar que a limitada comunicação entre os computadores desta época também limitava seus efeitos econômicos, o que foi mudando ao longo das décadas com o desenvolvimento de redes de conexão entre máquinas.

A primeira conexão por internet entre dois computadores ocorreu em 1969, ligando um ponto na Universidade da Califórnia em Los Angeles a outro na Universidade de Stanford

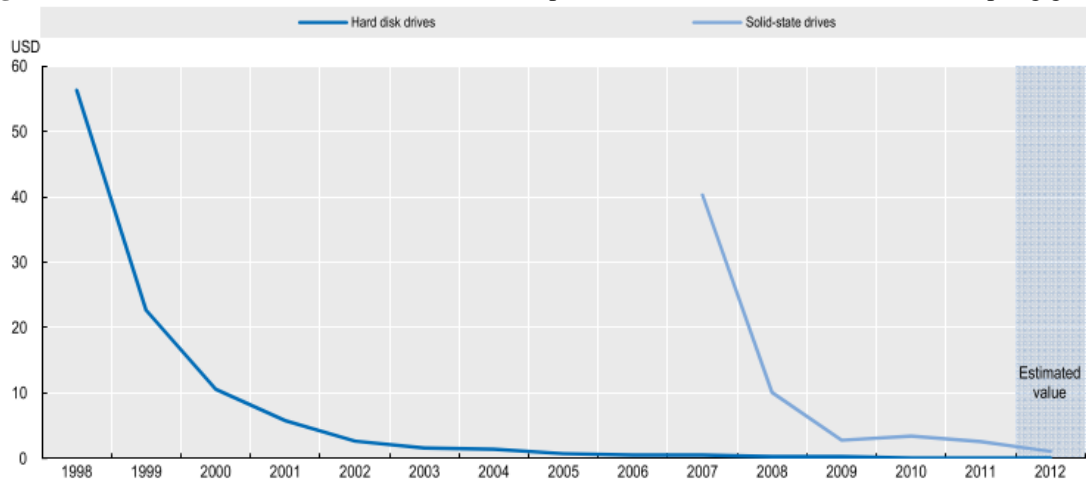
(SALUS, 1995). Deste ponto, então, gradativamente foram criando-se nodos de conexão, passando para 13 nodos em 1971, depois 23 nodos em 1972, 62 nodos em 1974 e 111 nodos em 1977 (HART; REED; BAR., 1992; O'NEIL, 1995). Até o final dos anos 1970 redes locais e internacionais foram criadas e conectadas via internet, sendo possível enviar dados por meio da comunicação de satélites (KITCHIN, 1998). Ao longo dos anos 1980 a infraestrutura das redes cresceu até que, finalmente, em 1992 é criada a *World Wide Web*, uma forma muito mais “amigável” aos usuários para acessar a internet (KITCHIN, 2014). Segundo Goldfarb e Tucker (2019), é somente neste período, junto com a privatização da tecnologia, que houve a rápida difusão da ferramenta para uso da sociedade. Conseqüentemente, observa-se o início de um período de expansão na intensidade do fluxo de informações que chegam aos agentes econômicos, alterando a dinâmica comportamental entre eles.

Cabe ressaltar, entretanto, que a grande mudança tecnológica em termos informacionais está, principalmente, ligada à nova infraestrutura que torna possível armazenar, pesquisar, copiar, filtrar, manipular e transmitir informação, como apontam Shaw, Shapiro e Varian (2000). Os autores fazem uma comparação lúdica: “Infrastructure is to information as a bottle is to wine: the technology is the packaging that allows the information to be delivered to end consumers”. (SHAW; SHAPIRO; VARIAN, 2000, p. 8).

Em termos econômicos, o que é realmente novo nesse contexto é nossa habilidade de manipular e acessar as informações, e não necessariamente o volume em termos quantitativos. A informação adquire diferentes características e definições de acordo com seus agrupamentos, guiada em alguns aspectos pelas mudanças tecnológicas, porém a aplicação das leis econômicas se mantém a mesma (SHAW; SHAPIRO; VARIAN, 2000). O que é necessário para compreender esta nova dinâmica da economia digital é, na verdade, a mudança na ênfase dada à análise (GOLDFARB; TUCKER, 2019).

Ao longo do período pós-expansão da internet nos anos 1990 e 2000, novas ferramentas e tecnologias de conexão foram desenvolvidas, como as redes locais de Wi-Fi e 3G, gerando uma expansão das plataformas digitais, a fim de atender às novas necessidades dos usuários. Navegadores, mecanismos de pesquisa, lojas online, redes sociais, entre outros são conseqüências desta evolução e diretamente responsáveis pelo crescimento na coleta e no uso de dados na economia (GOLDFARB; TUCKER, 2019). Em grande parte, a ascensão das plataformas digitais e seus novos modelos de negócios podem ser explicados tanto pelo ambiente tecnológico favorável à inovação, que reduziu os custos de coleta e armazenamento de dados (Figura 1), quanto pelas melhorias nas capacidades de programação e no desenvolvimento de *hardwares* de informação (YAMANA; NIREI; YAMANA, 2019).

Figura 1 – Custo médio de armazenamento de dados por consumidor (1998-2012) em dólares por gigabyte

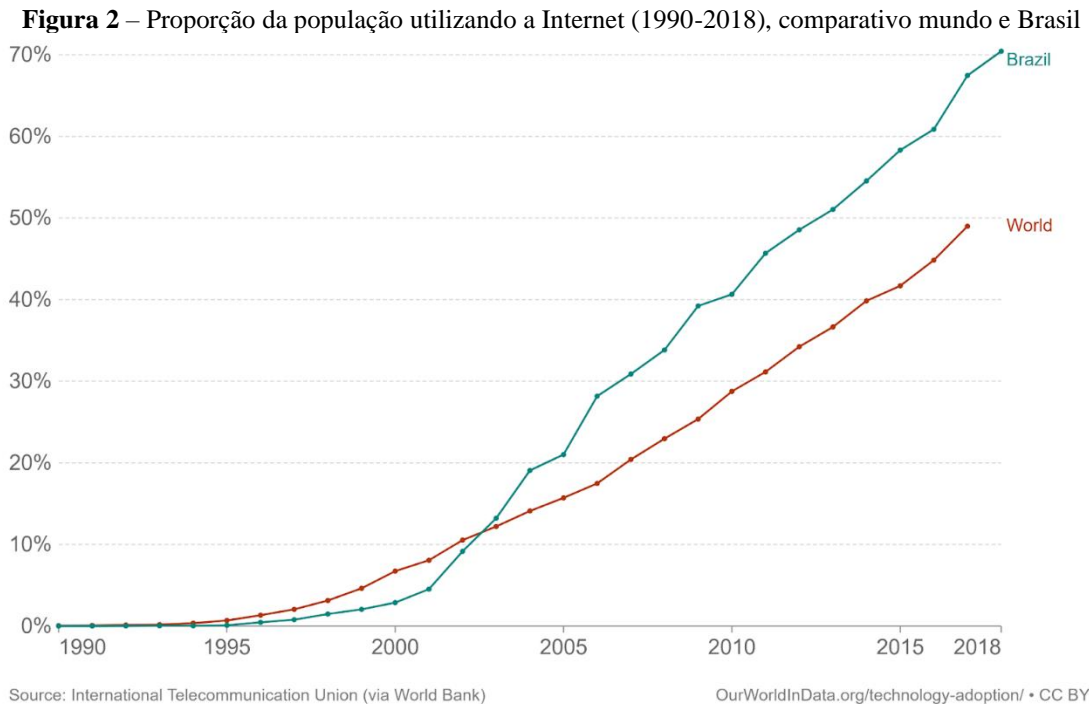


Note: Data for 1998-2011 are based on average prices of consumer-oriented drives (171 HDDs and 101 SSDs) from M. Komorowski (www.mkomo.com/cost-per-gigabyte), AnandTech (www.anandtech.com/tag/storage) and Tom's Hardware (www.tomshardware.com). The price estimate for SSD in 2012 is based on DeCarlo (2011) referring to Gartner.

Source: Based on Royal Pingdom blog, December 2011.

Fonte: OCDE (2015, p. 145)

Para se ter uma noção da popularização do uso da internet nos tempos atuais, dados de 2017 fornecidos pelo Banco Mundial registram uma grande evolução em relação ao uso da internet ao redor do mundo (Figura 2). Em 2000, período de plena expansão da ferramenta digital, apenas 6,7% da população mundial utilizava a internet. Uma década depois, em 2010, esse número saltou para 28,8%, chegando a quase a metade da população em 2017, com 49% dos indivíduos ao redor do mundo utilizando a internet. No Brasil, entre 2000 e 2019, a proporção do uso entre a população nacional avançou de 2,87% para 70,43%.



Fonte: Our World in Data (2018).

No gráfico, são considerados todos os indivíduos que utilizaram a internet nos últimos três meses anteriores ao momento da pesquisa, o que inclui o uso via computador, telefone móvel, máquinas de jogos, TV digital etc. Observa-se que os dados vão apenas até 2018, existindo ainda um grande potencial de aprofundamento da digitalização da economia no período recente da pandemia do Covid-19 iniciado em 2020, no qual diversos bens e serviços se viram forçados construir infraestrutura capaz de atender por meios digitais a fim de respeitar os distanciamentos sociais. Entretanto, para entender o comportamento histórico do uso da internet, os dados apresentados são suficientes para a construção do argumento nos próximos capítulos.

2.2 A EVOLUÇÃO DA INFORMAÇÃO DIGITAL ATÉ O *BIG DATA* E A FORMAÇÃO DA ECONOMIA *DATA-DRIVEN*

Assim como a Revolução Industrial teve grandes impactos em termos econômicos e sociais ao longo do século XVIII e XIX, muito se referencia o poder transformador da revolução informacional das últimas décadas, junto ao crescimento das tecnologias de informação e comunicação (TIC) e à reestruturação generalizada da economia global em relação a serviços e à tecnologia. Diferentemente dos períodos anteriores, no qual se dava grande importância aos setores primários da economia, na nova era este foco é direcionado ao uso das tecnologias de

informação (OJANPERÄ; GRAHAM; ZOOK, 2019; BRYNJOLFSSON; MCAFEE, 2016; SCHWAB, 2016).

Pensando no atual estágio de evolução tecnológica e na estruturação conceitual de uma economia digital, os termos “informação” e “dado” passam por algumas modificações em relação a seus significados. Pela definição mais específica de Setzer (2001, p. 1), antes de se tornar uma informação, um dado é “uma sequência de símbolos quantificados ou quantificáveis”, o que inclui textos (conjunto de símbolos de letras do alfabeto), fotos (conjunto de cores), figuras, sons etc. Os dados não necessariamente são ligados a meios digitais, existindo também em formato analógico, como sempre ocorreu dentro da experimentação científica, por exemplo. O termo “dado” foi primeiro utilizado em inglês (*data*) no século XVII e, ao longo do tempo, a palavra adquiriu um significado generalizado como sendo algo pré-analítico e pré-factual, que precede fatos, evidências, informações e conhecimento (KITCHIN, 2014).

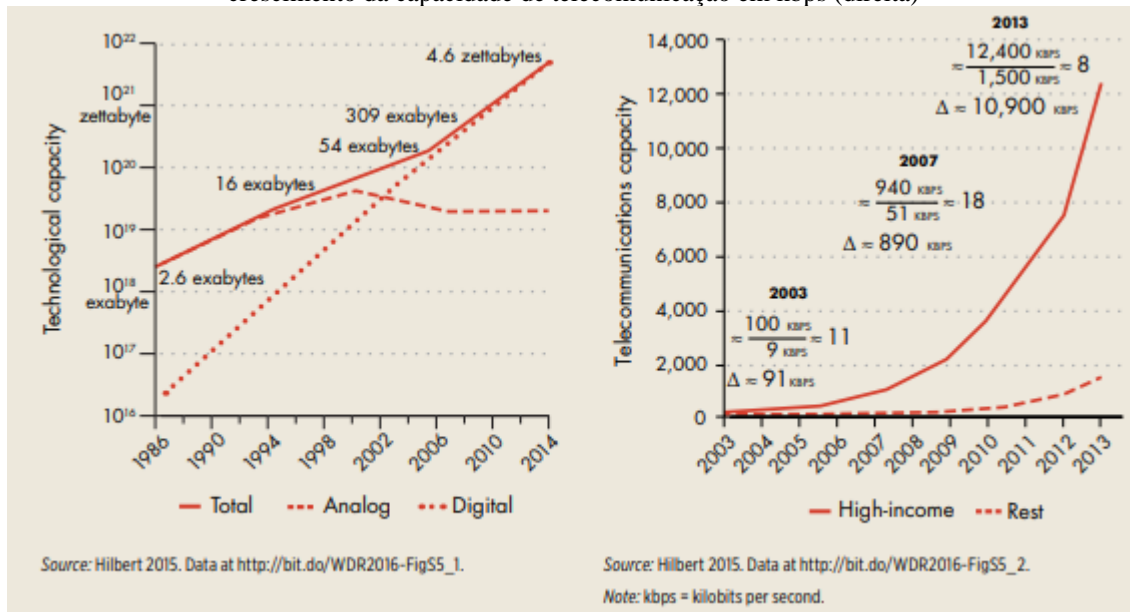
Com a mudança de contexto tecnológico, o termo adquire novo sentido como sendo uma coleção de elementos binários, no qual Setzer (2001) comenta: “O processamento de dados em um computador limita-se exclusivamente a manipulações estruturais dos mesmos, e é feito por meio de programas. Estes são sempre funções matemáticas, e, portanto, também são ‘dados’” (SETZER, 2001, p. 2).

Utilizando o entendimento de Kitchin (2014), de que “informação” é o acúmulo de dados associados entre si com algum grau de análise e correlação, no contexto da economia digital – o qual será utilizado neste trabalho – complementa-se com a simples e direta definição de Shaw, Shapiro e Varian (2000, p. 3) sobre o atual armazenamento de informação como “qualquer coisa que possa ser digitalizada ou codificada como um fluxo de *bits*”. Essa definição pode ser utilizada como um termo genérico para nos referirmos a o que é chamado hoje de “conjunto de dados” e que abrange qualquer tipo de registro armazenado no meio digital, estruturado ou não, como simples observações sobre objetos do mundo real ou virtual.

Com a transformação tecnológica, o desenvolvimento de infraestruturas de rede e dos mercados de produtos digitais, rapidamente diversas ferramentas eletrônicas foram adentrando o dia a dia de empresas e consumidores, como computadores e telefones móveis, tornando-se cada vez mais baratas, portáteis e rápidas, expandindo sua acessibilidade em uma escala global. Neste ritmo de transformação, Kitchin (2014) cita alguns dos principais fatores determinantes para o surgimento do fenômeno denominado *big data*: a criação de dispositivos e sistemas puramente digitais; o desenvolvimento da habilidade de acessar redes de comunicação e computadores em diversos tipos de ambientes, assim como em trânsito; avanços em estruturas

e sistemas de armazenamento e gerenciamento de dados (Figura 3); distribuição de armazenamentos de dados a custos acessíveis; e novas formas de análises de dados, que auxiliam na forma de lidar com a informação em abundância.

Figura 3 – Capacidade mundial de armazenamento em bytes entre meios analógicos e digitais (esquerda) e crescimento da capacidade de telecomunicação em kbps (direita)



Fonte: The World Bank (2016, p. 245).

Como resultado dessa aceleração da economia digital, naturalmente são produzidos dados em quantidade e velocidade explosivas. Os diversos atores da economia como a indústria, governo e universidades vêm acumulando uma quantidade massiva de dados cuja informação tem sido extraída a fim de gerar novos conhecimentos. A capacidade de armazenamento dos computadores vem aumentando cada vez mais, assim como se observa uma constante melhoria nos mecanismos de troca e transmissão de dados. Para ilustrar em números a evolução da magnitude desta nova dinâmica, segundo Eaton e Zikopoulos (2012), no ano 2000 cerca de 800 mil *petabytes* de dados estavam armazenados ao redor do mundo. Em outro levantamento de Short, Bohn e Baru (2011), os servidores mundiais processaram 9,57 *zettabytes* em informação, o que significa 12 *gigabytes* de informação diária por trabalhador ou 3 *terabytes* de informação anual por trabalhador. Em 2018, os usuários e as empresas ao redor do mundo geraram 33 *zettabytes* de informação, estimando-se que em 2025 este número salte para 175 *zettabytes* (NOVIKOV, 2020).

Diante do crescimento na quantidade de dados digitais disponíveis no mundo e no estabelecimento cada vez mais forte do *big data* dentro da sociedade, obviamente surge um grande interesse em relação aos potenciais valores comerciais que este fenômeno pode trazer

aos atores econômicos, o que serve como motivação para a estruturação da chamada economia *data-driven*, isto é, uma economia cuja decisões são baseadas e direcionadas por meio de dados. Como aponta Ciuriak (2019), é importante salientar que, antes da eclosão dos dados digitais em abundância, existiam processos de caráter *data-driven* no dia a dia da sociedade, como, por exemplo, a criação de indicadores da bolsa de valores pelo editor do *Wall Street Journal*, Charles Dow, e o estatístico Edward Jones, publicado em 1896, com o propósito de identificar padrões de comportamento entre os investidores. Outro exemplo são os cálculos das contas nacionais com o propósito de guiar as decisões de políticas públicas do governo. Até mesmo no ramo dos esportes existiam decisões *data-driven*, como a denominada estratégia “*money ball*”, que analisava as estatísticas dos jogadores de *baseball* e direcionava suas contribuições nas vitórias durante os jogos e, assim, seu valor no mercado.

Uma das maiores distinções entre a economia *data-driven* moderna e as aplicações de dados em antigas decisões econômicas está na eficiência da capacidade analítica sobre um grande volume de dados, isto é, análises que poderiam levar dias ou semanas agora levam apenas alguns minutos. Ciuriak (2019) explica:

The distinction between the modern data-driven economy and these examples of data-driven behavioural change lies primarily in the transition from gathering the low-hanging fruit of data analysis based on pattern that is accessible to the human mind to extraction of pattern from ‘big data’. The information in ‘big data’ is, almost by definition, not something that the human mind can access (to us, it is noise; the machine teases out the pattern) (CIURIK, 2019, p. 3).

Além desta melhoria na eficiência das análises, Yulinsky (2012) também cita o importante aprofundamento da autonomia na tomada de decisão, tendo em vista a possibilidade de implementação de processos de automação com o auxílio de ferramentas modernas de ciência de dados.

Entretanto, dando um passo atrás no raciocínio econômico, para entender a lógica por trás da economia *data-driven* também se deve compreender o porquê de os dados terem se tornado tão preciosos economicamente. Antes de tudo, a informação por si só, além de ser qualquer coisa passível de digitalização por meio de um conjunto de dados, também é um bem pela qual as pessoas estão dispostas a pagar, independente da sua fonte particular de valor (revistas, bancos de dados, filmes, músicas, softwares, entre outros) (SHAW; SHAPIRO; VARIAN, 2000). Desde os anos 1980, Knox e Castells (1988) argumentavam sobre a etapa do capitalismo da época, no qual o conhecimento seria o fator gerador das reestruturações econômicas, melhorando os produtos disponíveis no mercado, aumentando a eficiência

produtiva e a competitividade, entre outras profundas transformações no paradigma socioeconômico.

De fato, no atual estágio de evolução da economia da informação e diante do fenômeno do *big data*, presencia-se mais um avanço sobre essa lógica defendida por Knox e Castells, na qual os dados digitais servem ao propósito de auxiliar os agentes econômicos a monitorar e avaliar indicadores personalizados de comportamentos econômicos em tempo real, reduzir problemas de desperdício e fraudes e criar planejamentos sobre o mundo real de forma mais precisa. Especificamente para as empresas, o *big data* no contexto da economia *data-driven* também as auxiliam na criação e identificação de novos produtos, além de criar e direcionar mercados, com preços dinâmicos e maior produtividade (MANYIKA *et al.*, 2011; ZIKOPOULOS *et al.*, 2012).

O rápido desenvolvimento na complexidade das redes de computadores, do conceito do *Internet of Things* (IoT) e a computação em nuvem também foram fatores chaves na explosão do crescimento dos dados em quase todas as áreas da economia (JIN *et al.*, 2015). Os dados ganharam uma importância significativa a ponto de surgir mercados de dados digitais. As atividades relacionadas à coleta, consolidação e análise de dados se tornaram uma indústria multibilionária, derivada de diversos setores da economia como turismo, logística, saúde, imobiliário, e muitos outros. Com o crescimento dos meios digitais e da internet, a indústria se expandiu, com dados mais diversificados, novos produtos e serviços digitais e maiores possibilidades comerciais (KITCHIN, 2014).

Kitchin (2014) exemplifica como esses dados vendidos podem ser aplicados nos negócios:

Selling data to brokers has become a significant stream of revenue for many companies. For example, retailers often sell on data concerning transactions such as credit card details, customer purchases and store loyalty programmes, customer relationship management, and subscription information. Websites sell on clickstream data concerning how a person navigated through a website and the time spent on different pages. Likewise media companies, such as newspapers, radio and television stations, harvest the data contained within their content (e.g., news stories, adverts). Social media companies harvest the metadata and content of their users' interactions, using the data to construct consumer profiles and their own data products, or selling on the data to data brokers (KITCHIN, 2014, p. 42).

Com a explosão do uso de aparelhos digitais entre os diversos consumidores ao redor do mundo, produzindo dados, interagindo entre si e trocando informações, junto ao aperfeiçoamento nas ferramentas de manipulação e programação digital e o rápido declínio dos preços das tecnologias de informação, naturalmente surgem novos modelos de negócios com

estratégias e produtos *data-driven* mais sofisticadas, incluindo as modernas “plataformas digitais”, tão presentes em nosso atual cotidiano (YAMANA; NIREI; YAMANA, 2019). O crescimento dessas plataformas nos últimos anos foi monumental (Figura 4), o que os coloca na economia atual em uma posição de grande poder, em parte explicado pela detenção de grande volume de dados, além de fatores econômicos específicos, conforme Levin (2011) explica:

These and other internet platforms all take advantage of how the internet has lowered a range of economic costs: the cost of creating and distributing certain types of products and services, the cost of acquiring information about these goods, the cost of collecting and using data on consumer preferences and behavior. These changes have helped make internet platforms particularly dynamic and innovative, and inspired a great deal of economic research (LEVIN, 2019, p.1).

Figura 4 – Capitalização de mercado das maiores empresas *data-driven* (2019) em bilhões de dólares

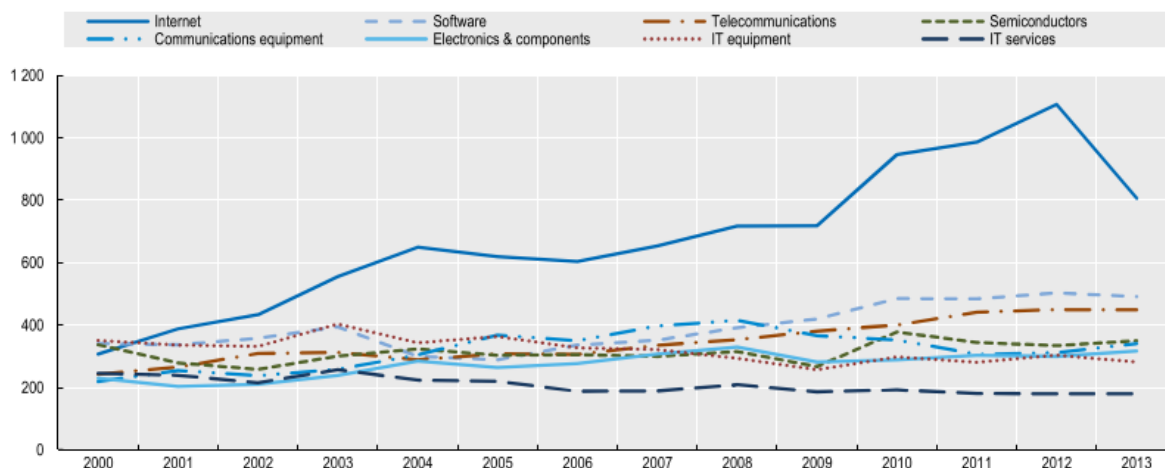
Company	Market Capitalization, 26 April 2019 USD billions
Microsoft	995
Apple	963
Amazon	960
Alphabet	888
Facebook	547
Netflix	164
Total	4,517
Memo: S&P500	23,770
Of which intangibles	21,390

Source: compiled from Internet sources

Fonte: Ciuriak (2019, p. 7).

O modelo de negócio das chamadas “empresas ponto com”, baseadas em plataformas digitais, está diretamente envolvido com a coleta e análise de um grande volume de dados fornecidos por seus próprios consumidores como usuários da internet. Com essa dinâmica de funcionamento e investindo fortemente em capital baseado em conhecimento (*knowledge-based capital* ou *KBC*, em inglês), essas empresas conseguem experimentar e inovar em termos de novos produtos e modelos de negócio de forma muito mais rápida e eficiente do que outros setores da indústria na economia (OCDE, 2015). Segundo dados levantados pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), entre as principais empresas no ramo da tecnologia da informação e comunicação (TIC), as “empresas ponto com” geraram, em média, mais de 1 milhão de dólares de receita por empregado em 2012 e mais que 800 mil dólares em 2013, enquanto as outras empresas do ramo geraram entre 200 mil (empresas de serviços de TI) a 500 mil dólares (empresas de software), conforme Figura 5:

Figura 5 – Média de receita por empregado entre as 250 maiores empresas de TIC (2000-2013) em um mil de dólar



Note: The presentation is based on averages for those firms reporting in 2000-13.

Sources: Based on OECD Information Technology database; compiled from annual reports, SEC filings and market financials, July 2014.

Fonte: OCDE (2015, p. 23).

A relevância econômica dos dados no meio digital está diretamente relacionada à digitalização dos bens e serviços, tendo em vista que a expansão das atividades econômicas e sociais realizadas por meio de plataformas digitais produz cada vez mais dados, em um processo de retroalimentação de informações, isto é, uma transação de compra online produz como resultado, além do dado da venda, uma informação nova sobre o indivíduo que a realiza – isto é, sua preferência. Embora aparentemente sem valor em um primeiro momento, os avanços em técnicas analíticas como inteligência artificial e *machine learning* permitiram a extração de informação valiosa destes dados disponíveis de forma eficiente, transformando em conhecimento estratégico aplicado aos negócios (CARRIÈRE-SWALLOW; HAKSAR, 2019).

Junto à emergência do *big data* como um fenômeno regular no cotidiano da sociedade, as modernas tecnologias e estratégias de negócio buscam novas formas de utilizar (e reutilizar) os dados digitais e a informação como uma fonte de extração de valor, em busca de abrir caminho para a criação de novas vantagens competitivas no mercado. Este processo envolve uma série de etapas e atividades conectadas entre si e com um fluxo lógico de agregação de valor, caracterizando o que Porter (1985) denominou de “cadeia de valor”. Rayport e Sviokla (1995) foram os primeiros a aplicar o conceito de Porter ao funcionamento de um sistema ligado à dinâmica de valoração da informação. No contexto do *big data*, existem diversas atividades na extração de valor dos dados e na criação de conhecimento de negócio, o que passa pelas etapas de geração, coleta, agregação, análise e, por fim, seu uso e monetização de fato (FAROUKI, 2020).

2.3. APLICAÇÕES DOS DADOS DIGITAIS NO SÉCULO XXI

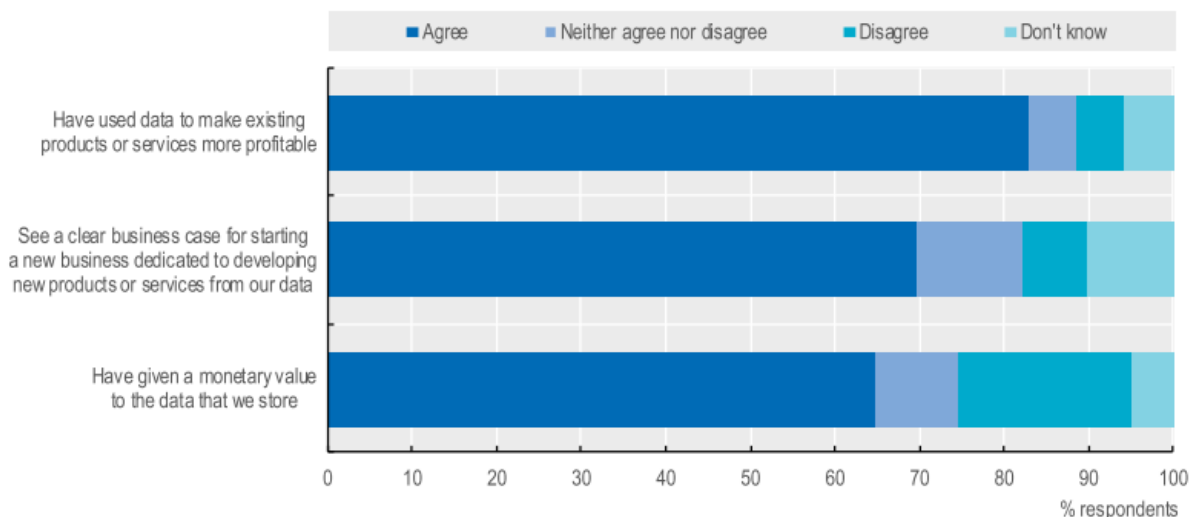
O novo contexto estabelecido pelos avanços tecnológicos e surgimento de novas ferramentas no cotidiano dos indivíduos em escala global criou uma gama de aplicações para o uso racional e eficiente dos dados a favor de determinados propósitos, o que inclui instituições governamentais, empresas privadas e os próprios consumidores. Os atores econômicos estão percebendo que a informação tem sido um fator estratégico de desenvolvimento, assim como o reconhecimento de novas necessidades de adaptações à realidade, seja por técnicas de manipulação de dados ou por novas estruturas jurídicas por parte do governo para conciliar e proteger essas informações de forma adequada. A coleta e análise de diversos conjuntos de informações e o desafio de extrair o máximo de valor e oportunidades dos dados, de forma ética e segura, ainda será um tópico em constante evolução por diversos anos.

Atualmente, considerando o meio dos negócios, as aplicações do *big data* para explorar vantagens de mercado são uma realidade em contínua expansão. Segundo relatório da OCDE (2013b), os dados vêm se tornando uma importante fonte de criação de valor e formação de capital baseado em conhecimento, fornecendo grandes melhorias em termos de produtividade. Ou ainda, a possibilidade de utilizar conjuntos de dados sobre padrões de comportamento dos consumidores para identificar novas tendências de demandas e tornar a previsibilidade de ajustes de produção e preços cada vez mais eficientes. Um estudo de Erik Brynjolfsson do Massachusetts Institute of Technology (MIT) realizou um levantamento que conclui que firmas que adotam decisões baseadas em dados atingem resultados em produtividade de 5% a 6% maiores do que seria esperado em outros tipos de investimento e uso de tecnologias de informação (BRYNJOLFSSON; HITT; KIM, 2011).

A divisão de pesquisa e análise do Economist Group que fornece serviços consultoria de análises e previsões econômicas, a Economist Intelligence Unit, conduziu uma pesquisa em 2015 na qual foi levantada a informação de que quase 90% dos executivos entrevistados afirmam que os dados estavam presentes na maioria das etapas do negócio; 83% dizem que suas empresas utilizavam conjuntos de dados para melhorar a lucratividade de seus produtos; 70% enxergavam potencial de criação de novos produtos ou serviços a partir de seus dados; por fim, mais de 60% verificava valor monetário e possibilidade de vendas com seus dados (Figura 6) (NGUYEN E PACZOS, 2020; THE ECONOMIST INTELLIGENCE UNIT, 2015).

Figura 6 — Resposta da pesquisa “*The Business of Data*”

Do you agree or disagree with the following statements about your organisation?



Note: Agree summarises “strongly agree” and “somewhat agree” responses. Disagree summarises “somewhat disagree” and “strongly disagree” responses. Based on a survey of 476 senior executives worldwide.

Fonte: OCDE (2020, p.15).

Além do uso das firmas, o *big data* também gera grande potencial de aproveitamento em eficiência no funcionamento das instituições governamentais e de políticas públicas. A melhoria na transparência da gestão também se torna um fator interessante neste contexto, pois podem publicar dados relevantes ao público por meio de plataformas online, permitindo também, junto com as novas tecnologias de comunicação e crescimento das redes sociais, estabelecer um papel mais proativo dos cidadãos sobre a política local (OCDE, 2013b). O terremoto que ocorreu no Haiti em 2010 é um exemplo de uso de tecnologias de comunicação e dados digitais de forma cooperativa para auxiliar no atendimento àqueles atingidos pelo desastre. Uma plataforma foi utilizada para mapear dados dos dispositivos móveis da população, estruturando um sistema para decodificar em tempo real dados de geolocalização e rotas de transmissão com o propósito de fornecer informações para os times de buscas e resgate em solo, melhorando a eficiência da ação e salvando mais vidas (APOLITICAL, 2017).

Outro exemplo interessante é o caso do “*Vlaamse Dienst voor Arbeidsbemiddeling*” (ou VDAB) na Bélgica. Trata-se de um serviço público de emprego de Flandres, que tem como objetivo apoiar os cidadãos a encontrarem oportunidades de trabalho, fornecendo um grande banco de dados online com informações de vagas de emprego. O serviço conta com algumas soluções *data-driven* como o “*job match*”, que recomenda vagas personalizadas que são baseadas nas características do perfil dos indivíduos e de acordo com suas competências. Outra iniciativa é o uso de modelos estatísticos preditivos, que utilizam dados para prever em quanto

tempo aproximadamente um desempregado conseguirá uma alocação no mercado de trabalho (BARBERO *et al.*, 2016).

É inegável que o uso do *big data* se tornou inevitável em todos os setores da sociedade, mas é importante observar os principais impactos dessa adoção dos dados como pilar na tomada de decisões. De acordo com o cientista político Steven Weber, embora dados sejam “baratos” e existam em abundância em qualquer localidade, suas ferramentas de análise (*data analytics*) são caras pois dependem de infraestrutura e mão de obra qualificada, o que pode acabar diminuindo o potencial de aproveitamento dos benefícios do *big data* em regiões que possuem deficiência desses fatores (WEBER, 2017).

Além disso, existem questões importantes de governança em relação aos dados, que também possuem ligação com a economia *data-driven*. Aaronson (2019) faz levantamento de alguns indicadores sobre como os diversos países lidam com seus dados e conclui que países mais ricos geralmente são mais abertos e possuem mais dados aplicados de forma eficaz para a resolução de problemas e criação de riqueza, ao mesmo tempo em que possuem melhores mecanismos de proteção de seus dados. Este tipo de desigualdade na governança de dados e uma maior fragilidade na segurança das informações em algumas localidades pode gerar algumas consequências éticas e econômicas, que podem desencadear uma reação em cadeia na confiança dos indivíduos sobre a economia *data-driven*.

2.4 NOVAS IMPLICAÇÕES ÉTICAS NO CONTEXTO DIGITAL

O contexto com novas estruturas tecnológicas e diferente interação com a sociedade necessariamente traz consigo uma série de implicações de poder, tendo em vista que essas ferramentas são utilizadas por determinados atores econômicos para benefício social, monetário e político. Apesar de parte dos efeitos do *big data* serem positivos, como a aceleração da inclusão digital, melhorias na transparência das ações por parte de instituições governamentais, a grande capacidade inovadora nos negócios e o melhor atendimento às necessidades imediatas dos consumidores, é importante observar as implicações da facilidade maior em captar informações de cada vez mais indivíduos e suas vidas. Especificamente no contexto dos dados, a preocupação é voltada para quais dados são gerados, como são criados, de que forma são compartilhados ou comercializados, quais os mecanismos de proteção são empregados e para quais fins são utilizados (KITCHIN, 2014). Conforme explicam Stucke e Ezrachi (2017):

Big data, sophisticated computer algorithms, and artificial intelligence are not inherently good or bad, but that doesn't mean their effects on society are neutral. Their nature depends on how firms employ them, how markets are structured, and whether firms' incentives are aligned with society's interests (STUCKE; EZRACHI, 2017, p. 2).

Uma das questões mais discutidas – e que terá grande enfoque neste trabalho – trata sobre a categoria de dados pessoais. Existem recorrentes debates sobre como as diversas formas de geração de dados sobre os indivíduos e as possibilidades de análises sobre essas informações podem impactar nos seus direitos de privacidade ou até mesmo sobre os direitos humanos, podendo ser utilizados como uma ferramenta de discriminação social conforme seu conjunto de características (GRAHAM, 2005). Trata-se de preocupações válidas, pois, com a grande adesão ao uso de plataformas digitais e de dispositivos móveis com diversos sensores de registro de informação dos usuários, praticamente qualquer tipo de ação individual é capaz de deixar um “rastro digital”. Além disso, o atual estágio de desenvolvimento das tecnologias de comunicação permite que os dados sejam facilmente criados e transferidos.

Alguns casos de usos indevidos de dados por parte de empresas ganharam atenção entre noticiários nos últimos anos, sendo um dos exemplos mais emblemáticos o escândalo do Facebook e a Cambridge Analytica, uma empresa britânica de consultoria política que realiza análises baseadas em dados para campanhas eleitorais. A empresa britânica comprou dados de milhões de usuários da rede social com o propósito de criar mecanismos para auxiliar na eleição de um dos candidatos americanos em 2016, Donald Trump. Essa estrutura analítica é baseada em um artigo da Universidade de Cambridge de 2013, que explicava que seria possível prever a personalidades de indivíduos através das páginas curtidas em seu perfil no Facebook – um tipo de análise denominada “*profiling*”. Um outro pesquisador da mesma universidade, Aleksandr Kogan, havia criado um teste de personalidade na rede social e alegava que teria propósitos acadêmicos. Com mais de 270 mil usuários que fizeram os testes, os dados de todos os contatos (ou “amigos”) de cada um dos respondentes também foram recolhidos – o que, de antemão, exemplifica uma externalidade sobre a privacidade dos indivíduos dentro das plataformas digitais –, estimando aproximadamente 87 milhões de usuários afetados.

Os dados, então, foram vendidos para a empresa de consultoria política e utilizados para alimentar um algoritmo que criaria grupos de perfis psicológicos para os usuários e forneceria conteúdos personalizados de cunho político de acordo com seu perfil (LAPOWSKY, 2019; STROUD, 2018). Em 2018, jornais como *The New York Times* e *The Observer* revelaram o escândalo, levando a Cambridge Analytica à falência e a divulgação de um pedido público de

desculpas por parte do CEO do Facebook, Mark Zuckerberg, além de ser convidado a se manifestar sobre o assunto perante o congresso norte-americano e o Parlamento Europeu.

No cenário brasileiro, um dos maiores exemplos é o caso do vazamento de dados da empresa Netshoes. Em janeiro de 2018, houve vazamento dos dados de quase dois milhões de contas com informações de usuários cadastrados no site da loja, devido a um ataque. O Ministério Público do Distrito Federal e Territórios (MPDFT) tratou o caso como “um dos maiores incidentes de segurança registrados no Brasil, colocando em risco os dados pessoais de milhares de consumidores” (MPDFT, 2018).

O resultado disso é que as pessoas estão começando a ficar mais cientes sobre os impactos da economia digital sobre si e alimentando uma maior proatividade em relação à proteção de seus direitos e privacidade. Da mesma forma, as empresas estão estimulando mais debates internos sobre valores éticos em relação aos consumidores e capacitando seus empregados sobre como devem lidar com as informações sensíveis no meio corporativo. Governos de diversos países ao redor do mundo vêm se preparando para lidar com as questões éticas ligadas ao tratamento dos dados digitais, criando novas legislações específicas para o fenômeno do *big data*.

A Europa é a região mais avançada em termos de estudos e leis sobre proteção de dados, principalmente entre os países da União Europeia. Desde os anos 90 o Parlamento Europeu discutia sobre o tratamento de dados pessoais, mas a grande evolução veio com a *General Data Protection Regulation* (GDPR), estabelecida em 2016. No estado da Califórnia nos EUA existe o *California Consumer Privacy Act of 2018* (CCPA), que entrou em vigor no início de 2020. No Canadá está estabelecida a *Personal Information Protection and Electronic Documents Act* (PIPEDA), existente desde 2000. Na Ásia, Oceânia e América Latina, países como Japão, Nova Zelândia e Argentina também possuem legislações bem estabelecidas sobre a questão.

Recentemente, em setembro de 2020, o Brasil colocou em vigor a sua Lei Geral de Proteção de Dados (LGDP), inspirada nos moldes da GDPR na Europa, que vale para todas as empresas que manipulam dados dos cidadãos brasileiros em território brasileiro. Com a nova lei, eventos de vazamentos de dados que eram ocultados pelas empresas de forma recorrente agora possuem uma atenção especial, exigindo que ocorra uma comunicação aos usuários e à Autoridade Nacional de Proteção de Dados, além do estabelecimento de um plano de ação claro para evitar futuros novos vazamentos.

3 DEFINIÇÃO DE ESCOPO CONCEITUAL

Uma vez colocado o objeto de estudo em perspectiva histórica, este capítulo buscará delimitar o escopo conceitual a ser tratado ao longo do trabalho, a fim de deixar claro quais são as definições utilizadas e até qual nível de complexidade será tratado, sempre se observando sob o contexto da economia digitalizada. Este capítulo se torna necessário pois dados e informações, por exemplo, comumente são utilizados de forma “grosseira” como sinônimos, entretanto o seu entendimento dentro na visão econômica se dá de formas distintas. Ao mesmo tempo, sob a perspectiva da área de pesquisa de informática, a definição dos dados digitais (e do *big data*) pode chegar a um nível de complexidade que vai além das questões que serão discutidas para tratar sobre as possíveis falhas de mercado que os circundam, sendo, portanto, necessário especificar quais significados e classificações serão utilizados.

A mesma delimitação de escopo também deve ser utilizada a fim de compreender de que forma os dados adquirem valor dentro do funcionamento da nova dinâmica da economia digital, chegando à definição da economia *data-driven*. Por se tratar de um tema complexo e relativamente recente, os conceitos estão em constante evolução e ainda possui pontos de debate dentro do meio acadêmico. Portanto, alguns dos termos e significados a serem utilizados possui referência à visão de alguns autores e instituições que se debruçaram de forma profunda em pesquisas sobre os tópicos estudados.

3.1 CARACTERIZAÇÃO DOS DADOS DIGITAIS

Para compreender de forma completa a força que os dados digitais adquiriram no novo contexto socioeconômico construído até aqui é importante abordar com o devido detalhe o que abarca a definição do objeto de estudo, quais são suas classificações existentes e em quais contextos específicos eles são aplicados. Trata-se do elemento principal e que guia a economia digital e a influência econômica do *big data* sob a ótica das falhas de mercado.

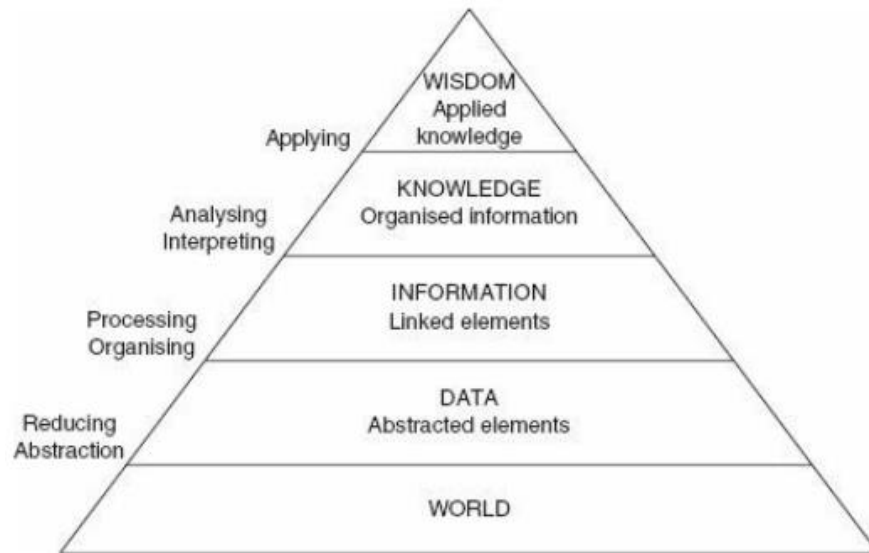
3.1.1 Definição

Algumas particularidades dos dados acabam direcionando seus efeitos econômicos, justamente devido às suas especificidades de comportamento. Portanto, é importante diferenciar conceitualmente os dados digitais de outros conceitos pré-existentes na teoria econômica, como a “informação” e o “conhecimento”, a fim de deixar claro como e onde as ideias da “economia

do conhecimento”, “economia da informação” e “economia *data-driven*” transitam na realidade e como se dá a diferença na criação de valor por meio delas.

Segundo Kitchin (2014), a etimologia da palavra “dado” deriva do latim “*dare*”, o que significa o ato de “dar”. Nesse sentido, o dado é um extrato de um fenômeno (e fornecido por este), que pode ser medido e registrado de formas diversas. As unidades de dados são aquelas selecionadas e colhidas de uma grande soma de potenciais outros dados (KITCHIN; DODGE 2011). Trata-se de um dos ingredientes a serem integrados no encadeamento de informações que formam uma ideia, sendo um material mais “bruto”, ainda desorganizado e sem uma estrutura inerente. É a partir deste encadeamento lógico e acumulativo sobre o conceito de dados que se pode definir “informação” e “conhecimento”. Porém, a forma que se entende o termo depende do contexto e da perspectiva nos quais está inserido, não sendo uma expressão nova do século XX, conforme explicado no Capítulo 2. Floridi (2008) explica alguns pontos de vista para entender os dados: da perspectiva epistêmica, por exemplo, os dados são um conjunto de fatos, sendo também informação do ponto de vista informacional. Do ponto de vista computacional – e na abordagem utilizada dentro da perspectiva da economia digital – os dados são uma coleção de elementos binários que podem ser processados e transmitidos eletronicamente.

Resumidamente, no escopo conceitual deste trabalho serão considerados dado, informação, conhecimento e sabedoria como partes de uma ordem sequencial (ou mesmo hierárquica) na qual o dado precede a informação, que por sua vez precede o conhecimento e a sabedoria (ADLER 1986; WEINBERGER, 2011). Esta estrutura conceitual pode ser ilustrada por meio de uma pirâmide (Figura 7), na qual cada camada é diferenciada pela execução de algum tipo de processo (simplificação, abstração, organização, análise, interpretação etc.), de forma que agregue significado e valor sobre o mundo ou fenômeno do qual foi extraído (KITCHIN, 2014).

Figura 7 – Pirâmide do conhecimento

Fonte: Kitchin (2014).

Esse processo de transformação do dado até o topo da pirâmide do conhecimento pode ser interpretado como uma espécie de “destilação”. Conforme Hey (2007) explica, uma grande quantidade de dados pode ser comprimida em uma pequena porção de informação. Para se entender o comportamento comercial de uma empresa, por exemplo, são necessários diversos dados sobre seus custos, vendas, capital etc., até finalmente chegar a determinada conclusão sobre sua posição de mercado.

É importante apontar que os dados e seus processos transformadores em conhecimento não existem de forma independente de um contexto e das pessoas que estão inseridas no mundo do qual a informação é extraída. O atual cenário do *big data* foi possível devido à construção de todo um processo de desenvolvimento tecnológico e à mudança do comportamento social por parte dos consumidores, empresas, governo e universidades. Como consequência dessa explosão no volume de dados produzidos e armazenados podem existir situações nas quais se tem muitos dados, mas pouca capacidade de extrair valor devido à deficiência ferramental (OCDE, 2013b; UBALDI, 2013). Da mesma forma, o excesso de informação pode não ser capaz de gerar conhecimento, conforme o economista vencedor do prêmio Nobel descreve “*a wealth of information creates a poverty of attention*” (SHAPIRO; VARIAN, 2000). A diversidade de procedimentos modernos de análise de um grande volume de dados e suas capacidades de transformação em informação ou conhecimento de forma mais eficiente do que antigamente decorre justamente de uma mudança de paradigma socioeconômico e tecnológico no qual a sociedade está vivendo.

Esse conjunto de técnicas e ferramentas que são utilizadas para transformar dados pouco estruturados em insumos de informação, conforme o processo sequencial mostrado na pirâmide do conhecimento (Figura 7), são denominados de “*data analytics*” (OCDE, 2013a). É importante entender quais são esses procedimentos disponíveis sobre a manipulação dos dados, pois são bastante utilizados por atores econômicos como empresas e governo a fim de extrair o contexto, organização e estruturação aos quais os dados estão relacionados, além de correlações sobre a realidade e possivelmente factuais (MERELLI; RASETTI, 2013). Utilizando como referência o artigo da OCDE (2013a), algumas dessas principais ferramentas de *data analytics* são referenciadas entre:

- ***Data mining***: É um conjunto de técnicas estatísticas (análise por *cluster*, classificação e regressão) em conjunto com métodos de gerenciamento (bancos de dados) e pré-processamento (filtragem e limpeza) de dados, a fim de realizar uma “mineração” preliminar de informações mais relevantes para o usuário.
- ***Profiling***: Utiliza a análise de dados para a construção de perfis ou classificação de entidades em determinados perfis de acordo com atributos específicos de cada grupo. É bastante utilizado para identificação de informações pessoais de determinado usuário, cálculos de *score* de crédito, discriminação de preços e para propagandas direcionadas (também conhecido como “*remarketing*”).
- ***Machine or statistical learning (ML)***: Trata-se de um campo de estudo na área de Inteligência Artificial (IA). Utiliza algoritmos que permitem à máquina “aprender” a executar determinadas tarefas, enquanto retroalimenta a melhoria de sua performance a cada novo conjunto de dados que é analisado (DUDA; HART; STORK, 2000; RUSSEL; NORVIG, 2009, HASTIE; TIBSHIRANI; FRIEDMAN 2011; JAMES *et al.*, 2013). Assim como o *Profiling*, também é de extremo interesse para entender algumas dinâmicas econômicas em relação ao comportamento sobre dados pessoais, conforme é tratado neste capítulo.

É importante notar também que, da mesma forma que um conjunto de dados pode explicar um fenômeno do mundo real, o processo inverso também é possível de ser realizado, isto é, com as ferramentas digitais atuais é possível capturar e transformar um evento do mundo real em dados quantificáveis (OCDE, 2019; 2013b). Este processo é denominado de

“*datafication*”, uma junção das palavras em inglês “*data*” e “*quantification*” (HEY, 2004; BERTOLUCCI, 2013; MAYER-SCHONBERGER; CUKIER, 2013), e pode ser extremamente útil para a tabulação e posterior análise de determinado fenômeno com uso de ferramentas de ciência de dados. O uso deste tipo de geração e análise de informações quantificáveis e armazenáveis ao longo dos anos e em diferentes localidades desempenha um papel muito importante no acompanhamento, registro e compreensão sobre a ocorrência dos diversos eventos históricos dentro da sociedade e do meio nos quais estão inseridos.

3.1.2 Classificação

No meio acadêmico não existe exatamente uma classificação universal que abarque todos os tipos de dados digitais, pois essas características dependem do contexto e premissas consideradas na análise sobre o seu uso. Entretanto, algumas instituições de regulamentação governamentais interessadas em criar políticas públicas sobre a governança de informações na economia digital precisam criar um *framework* conceitual sobre o objeto a ser regulamentado, a fim de limitar o escopo das regras estabelecidas e deixar claro quando elas são aplicáveis ou não. Sobre os dados pessoais, por exemplo, a *Swedish Board of Trade* (2015) utiliza uma definição desse tipo de dado em categorias que são relevantes para empresa e outras que são importantes para transferência de dados internacionais, diferenciando entre dados “voluntários”, “observados” e “inferidos”. Alternativamente, o Departamento de Comércio dos Estados Unidos classifica quatro tipos de dados de acordo com seu fluxo entre fronteiras e como influencia as estatísticas econômicas. Na União Europeia, com o Regulamento Geral sobre a Proteção de Dados (RGPD ou, em inglês, *GDPR*), os agentes econômicos recorreram às classificações da ISO/IEC de 1994 em “*Data flow, data categories and data use*”, que discute quatro diferentes classificações para os dados entre dados de conteúdo do usuário (informações biométricas e de contato), dados derivados (que identifique usuários finais ou organizações), dados por provedores de serviços na nuvem e dados contábeis (dados de pagamentos ou administrativos) (OCDE, 2013a).

Neste trabalho, será utilizada uma classificação simplificada, baseado em definidos da OCDE, e que varia de acordo com a posse ou direito de uso do dado, sua fonte e o método de coleta ou geração. São elas:

- **Quanto à posse ou direito de uso dos dados:** (a) de propriedade definida, na qual possui proteção via propriedade intelectual ou direitos similares, podendo ser tanto de

posse individual ou organizacional; (b) de domínio público ou *open data*, no qual o dado é disponível publicamente e livre para o uso;

- **Quanto à fonte do dado:** (a) dados pessoais, que se trata de qualquer dado que permite a identificação do sujeito da informação; (b) dados organizacionais, que descreve a identificação de organizações;
- **Quanto ao método de coleta ou geração do dado:** (a) dados criados por usuários; (b) dados gerados por meio de máquinas, como aqueles de comunicação entre máquinas (M2M) ou os resultantes de IoT (coletados de sensores).

É importante notar que as classificações não são mutuamente excludentes, sendo possível, portanto, classificar um conjunto de dados em mais de uma categoria, dependendo das circunstâncias. Além disso, serão utilizadas neste trabalho apenas as classificações relevantes no tocante ao tema a ser tratado, existindo, portanto, diversas outras características técnicas classificáveis sobre os dados e que fogem do escopo da análise a ser abordada nos próximos capítulos.

3.2 CARACTERIZAÇÃO DO *BIG DATA*

O primeiro uso do termo “*big data*” foi realizado por John Mashey, Cientista Chefe na Silicon Graphics, com o propósito de se referir à manipulação de um grande conjunto de dados (KITCHIN, 2014). Ao longo dos últimos anos diversas definições surgiram para o termo “*big data*”, conforme Cavanillas, Curry e Wahlster (2016) compilaram na Figura 8, e somente após meados de 2013 o termo ganhou grande popularidade nos noticiários e meio empresarial. Embora as diversas descrições não estejam incorretas, a primeira e mais referenciada definição vem de Doug Laney, que explica: “*Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization*” (LANEY, 2001). Sua conceitualização possui três dimensões sobre o *big data* (o que justifica o título de seu relatório de pesquisa não publicado, “*3D data management*”): volume, velocidade e variedade.

Figura 8 – Definições de “big data”

Big data definition	Source
“Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”	Laney (2001), Manyika et al. (2011)
“When the size of the data itself becomes part of the problem and traditional techniques for working with data run out of steam”	Loukides (2010)
Big Data is “data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time”	Jacobs (2009)
“Big Data technologies [are] a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis”	IDC (2011)
“The term for a collection of datasets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications”	Wikipedia (2014)
“A collection of large and complex data sets which can be processed only with difficulty by using on-hand database management tools”	Mike 2.0 (2014)
“Big Data is a term encompassing the use of techniques to capture, process, analyse and visualize potentially large datasets in a reasonable timeframe not accessible to standard IT technologies.” By extension, the platform, tools and software used for this purpose are collectively called “Big Data technologies”	NESSI (2012)
“Big data can mean big volume, big velocity, or big variety”	Stonebraker (2012)

Fonte: Cavanillas, Curry e Wahlster (2016, p. 31).

É importante observar que a grande quantidade de dados por si não é novidade na sociedade, tendo em vista o acúmulo de informações ao longo dos séculos em registros analógicos (livros, pinturas, fotografias, anotações etc.). Nas instituições governamentais, na indústria e nas universidades se utilizava uma grande quantidade de dados para extrair conhecimento, seja por meio de estatísticas sobre contas nacionais ou pesquisas científicas de qualquer ramo dentro da academia. Entretanto, o que o *Big Data* traz de novo é a maior complexidade e escala de dados, o que tornaria impossível a extração de informações sem novas ferramentas adequadas de análise. Essas novas características são resumidas pelos chamados três Vs, conforme explicam Cavanillas, Curry e Wahlster (2016):

- **Volume:** trata-se da quantidade de dados em larga escala;
- **Velocidade:** trata-se da velocidade dos dados, produzidos em tempo real;
- **Variedade:** trata-se da variedade de tipos e fontes de dados, com diferentes formatos e significados.

A nova dinâmica de funcionamento dos dados é sustentada pelo novo cenário econômico e tecnológico, com melhorias no meio da computação, formas alternativas de armazenamento de dados, barateamento de tecnologias e expansão de sensores que captam e transmitem dados em tempo real. Esse contexto exige, também, novas formas de processar, explorar e extrair valor de uma quantidade massiva de informações, que é de onde vem a necessidade de uso das ferramentas de *data analytics* (conforme citado no tópico 3.1.1.).

Com o desenvolvimento e maturação do meio do *big data*, novos Vs foram surgindo para acrescentar conceitualmente à definição do termo, como as características “veracidade” e “valor”. A veracidade diz a respeito da qualidade dos registros e a capacidade de verificar a acurácia da informação, a fim de evitar conclusões erradas extraídas dos dados (PFEIFFER, 2019). O valor é a característica definitiva do dado no contexto da economia digital que vem ganhando maior atenção por parte da sociedade e terá papel essencial ao longo da análise do trabalho, pois é de onde se extrai o benefício econômico do *big data*. Atualmente, as empresas possuem as ferramentas mais eficientes para coletar, armazenar, analisar e monetizar os dados, em busca de uma vantagem econômica na dinâmica da economia digital, chegando a serem considerados fatores de produção.

As atividades econômicas dentro da dinâmica do *big data* são diversas, envolvendo as várias etapas de seu funcionamento, como a coleta, armazenamento, análise e uso dos dados. A junção de todas as fases gera o que alguns autores chamam de “cadeia de valor dos dados” (“*data value chain*”) e “cadeia de valor do *big data*” (“*big data value chain*”), conceito que será desenvolvido mais à frente do trabalho.

3.3 CARACTERIZAÇÃO DA ECONOMIA *DATA-DRIVEN*

O conhecimento sempre teve um papel importante no desenvolvimento econômico e social, seja por meio de transmissão de informação com fins de registros históricos ou para criar métricas de acompanhamento das atividades da economia. No meio acadêmico as atividades guiadas por dados vêm de longa data para entender os fenômenos do mundo. Conforme desenvolvido no segundo capítulo, o mundo passou por diversas transformações e melhorias tecnológicas, o que permitiu uma maior produção e acesso a dados, tanto em volume como em qualidade. Esse fenômeno de “*datafication*” torna possível criar novas potencialidades de compreender e manipular as mudanças que ocorrem na realidade social, o que obviamente terá fortes impactos sobre a economia e o modo de viver da sociedade.

A explosão no uso dos dados no contexto do *big data* foi capaz de criar um ciclo virtuoso de análise e aplicação de dados em diferentes frentes da economia, criando cada vez mais fontes de dados e levando a análises mais complexas. Mais do que nunca as firmas estão investindo fortemente em produtos e serviços capazes de coletar informações sobre os consumidores e o mercado como um todo. Um estudo internacional realizado pela *European Patent Office* (EPO) analisou padrões de tendências em pedidos de registro de patentes relacionados a objetos da categoria “*smart connected objects*”, que está diretamente ligado a fatores da economia *data-driven*, como a internet das coisas (ou IoT), 5G, automatizações, inteligência artificial, *machine learning* etc. O estudo demonstra que, de 2010 a 2018, os pedidos de patentes para este tipo de categoria citada cresceram a uma média de 20% ao ano, um crescimento quase cinco vezes maior que aquele observado para outras categorias de patentes desde 2010 (4,2%). (MÉNIÈRE *et al.*, 2020).

Tanta criação de dados, entretanto, de nada adianta se não souber extrair seu valor de forma eficiente. Kennedy (2014) faz um comparativo sobre o potencial inexplorado da quantidade massiva de dados do *big data* com a existência de petróleo embaixo do solo.

For millennia, massive amounts of oil and natural gas lay trapped within shale deposits underneath the United States, making no contribution to economic growth. Within the last decade, new technology has allowed us to exploit these previously inaccessible resources (...). Big Data, as a resource, presents similar opportunities (...). Big Data has the potential to deliver large economic gains. The central challenge for public and private sector leaders is to apply this resource to the large variety of problems that now confront us while minimizing the relatively manageable risks associated with the greater availability of data (KENNEDY, 2014, p. 9).

Uma pesquisa realizada pela empresa de consultoria global McKinsey (2018) sugere que o aproveitamento e extração de valor monetário dos dados vem se tornando um fator cada vez mais importante no aumento da receita das empresas. Segundo o levantamento, a monetização de dados contribui em 11% ou mais no total da receita de 32% das empresas de alta performance avaliadas (isto é, aquelas que possuem um crescimento anual de 10% ou mais nos últimos três anos), e em 9% da receita total de todas as outras categorias de empresas.

Portanto, as empresas, em busca da melhor forma de conseguir gerar *insights* para o meio dos negócios, enxergaram na comunidade científica um papel crucial neste sentido, pois foi dali que se estruturaram ferramentas robustas o suficiente para facilitar e melhorar as análises baseadas em um grande volume de dados, o que seria então denominado “*data science*” (HEY; TANSLEY; TOLLE, 2009). Uma vez que dados se tornaram um novo fator de produção das firmas, da mesma forma que capital humano e ativos tangíveis, também são necessárias

peessoas capacitadas o suficiente para aplicar as novas técnicas de análise, gerando uma maior procura por cientistas de dados. Em resumo: seguindo a definição de uma das maiores empresas de banco de dados do mundo, a Oracle, a ciência de dados (ou “*data science*”) “combina múltiplas áreas de conhecimento, incluindo estatística, métodos científicos, inteligência artificial e análise dados, a fim de extrair valor dos dados” (ORACLE, 2021). Portanto, aqueles que praticam ciência de dados são cientistas de dados e combinam uma série de habilidades multidisciplinares para aplicar na análise de dados coletados de diversas fontes e sensores diferentes e gerar *insights* para os negócios.

3.3.1 Modelos de negócio: *data-enhanced* vs *data-enabled*

As novas ferramentas analíticas permitiram expandir o escopo e a escala do uso dos dados dentro da economia como um todo, abrindo espaço para a criação de novos modelos de negócios nos últimos anos. Os dados podem exercer diversos papéis como por exemplo auxiliar na coordenação de operações de negócio (*supply chain*), fornecer insumos para os processos decisórios e permitir a melhoria ou criação de novos bens ou serviços (NGUYEN; PACZOS, 2020). De modo geral, utilizando a taxonomia de Nguyen e Paczos (2020), pode-se classificar os modelos de negócio na economia *data-driven* em duas principais classes:

- **Modelo de negócio *data-enhanced*:** aqueles nos quais a aplicação do uso racional dos dados não altera ou determina o núcleo principal de seu modelo de negócio pré-existente (considerando aquele que determine mais de 50% da receita total), sendo uma nova maneira de agregar valor incremental;
- **Modelo de negócio *data-enabled*:** negócios que podem ser vistos como totalmente digitais, ou seja, toda sua estratégia de negócio é baseada em dados e sua existência está condicionada a um contexto de grande disponibilidade de dados e avançadas ferramentas de análise de dados (exemplos: Uber, Amazon, Twitter, Facebook, entre outros).

Em ambos modelos de negócio, a pergunta que guia seu funcionamento depende do propósito do que será oferecido, para quais grupos de consumidores, como se pretende extrair valor econômico e por qual preço. Porém, a escolha de qual o modelo de negócio é mais

adequado depende do quão dependente de dados seu produto ou serviço será para criar condições favoráveis de receita.

Uma abordagem que Kennedy (2014) utiliza para diferenciar a adoção dos modelos de negócio também depende de um determinado “problema” a ser resolvido. Se uma empresa pretende melhorar sua performance, pode adotar um modelo *data-enhanced* e realizar um levantamento de dados internos para identificar ou resolver um gargalo operacional, como por exemplo dificuldade na fidelização do consumidor, decisões ruins de negócio e uma administração falha de inventário ou estoque. A empresa de logística United Parcel Service (UPS) conseguiu aumentar sua receita e lucro por meio da coleta de informações sobre a atuação de sua frota de veículos, incluindo sua localização e performance. Por outro lado, as empresas que buscam desenvolver novos produtos ou serviços devem identificar quais são os problemas não resolvidos entre seus consumidores, abrindo espaço para a adoção de um modelo de negócio *data-enabled* e determinar quais serão os dados necessários para responder às dores dos clientes (KENNEDY, 2014).

Por fim, no panorama dos modelos de negócio *data-driven*, Nguyen e Paczos (2020) estabelecem quatro principais categorias (Figura 9): (i) categoria 1: Venda e licenciamento de dados (agregados ou não); (ii) categoria 2: Desenvolvimento e venda de novos produtos ligados a dados; (iii) categoria 3: Uso de dados para melhorar produtos preexistentes; (iv) categoria 4: Uso de dados para melhorar processos produtivos ou a eficiência do negócio. Essas categorias variam de acordo com a proporção de receita derivada da monetização de dados.

Figura 9 – O uso de dados: modelos de negócio *data-enabled* vs *data-enhanced*



Fonte: Nguyen e Paczos (2020, p. 16).

3.3.2 Plataformas digitais

Neste trabalho, as empresas que adotam o modelo de negócio *data-enabled* terão um enfoque maior, pois, devido às suas características altamente dependentes de dados para operacionalização, acabam gerando implicações mais específicas no contexto da economia digital. Uma das particularidades da maioria das chamadas “plataformas digitais” é que elas atuam sob um modelo de mercado de dois lados (M2L) ou mercado de múltiplos lados

(PFEIFFER, 2019). Segundo definição de Tirole (2017), este tipo de mercado permite que um intermediário atue como facilitador na interação entre vendedores e compradores, isto é, os próprios usuários são aqueles que ofertam e demandam. Exemplos desses intermediários podem ser referenciados pelas atuais plataformas digitais no mercado como Uber, Airbnb, eBay, Google etc. Neste contexto, além de fazer a intermediação entre duas ou mais partes, as plataformas também fornecem uma interface tecnológica que permite esta interação (TIROLE, 2017).

Observa-se que além do fato da maior parte da indústria na economia digital operar como plataformas sob o modelo de mercado de dois ou múltiplos lados, também possuem em comum o fato de darem uma grande importância estratégica aos dados dos usuários. Conforme classificação de Nguyen e Paczos (2020) sobre os modelos de negócio *data-enabled*, essas plataformas agregam valor ao seu negócio por meio da coleta e análise de dados de seus usuários, a fim de melhorar seu desempenho econômico e tornar a interface que possibilita a intermediação entre seus usuários cada vez mais eficiente. Essa importância é explicada pelo fato de a remuneração das plataformas serem oriundas dos dois lados do mercado que são intermediados, ou seja, empresas de produtos e serviços de cartão de crédito, por exemplo, precisam cobrar pagamento do usuário que utiliza o serviço bem como também do fornecedor. Neste tipo de dinâmica também é comum existir a atuação dos anunciantes, que utilizam o alcance da plataforma para fazer propaganda de produtos aos usuários.

Uma característica importante da maioria das plataformas digitais em mercado de dois ou múltiplos lados e que oferecem produtos ou serviços “grátis”, muitas vezes em troca da criação de um cadastro – como, por exemplo, redes sociais –, na verdade utilizam os dados pessoais dos usuários como moeda de troca. Isto é, embora muitas vezes não exista uma transação monetária entre o usuário e a plataforma para se utilizar “gratuitamente” os serviços oferecidos pela interface, os dados do usuário são coletados, analisados e monetizados pelo modelo de negócio (explicitado pela Figura 10).

Figura 10 – Dados de *input* e *output* para plataformas digitais

INPUT	OUTPUT
Product/Services (seller)	New data product
Personal data (buyer)	Pattern/trends/correlations for sales/marketing/product development
Spatial-temporal data (buyer)	Customized offers
Data collected by sensors (IoT)	Customized advertising
Community information (Social Networks)	Community suggestions
Transactional data	Operational data

Fonte: Visconti, Larocca e Marconi (2017, p. 10).

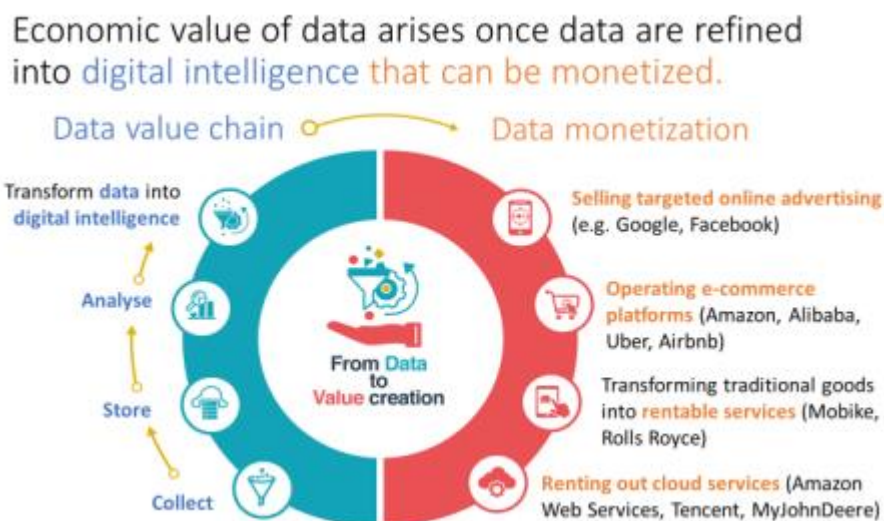
Observa-se que, utilizando-se dos dados pessoais como *input*, alguns dos resultados deste uso por parte das plataformas digitais é a customização de ofertas e propagandas e o atendimento às características específicas do consumidor. Este e outros comportamentos geram uma série de implicações econômicas como externalidades, assimetria de informação e poder de mercado, adentrando também no escopo ético da discussão sobre privacidade do usuário, e são questões que serão abordadas mais a fundo no capítulo seguinte.

3.3.3 Cadeia de valor dos dados (“*Data value chain*” e “*Big data value chain*”)

Para entender de que forma se dá a monetização dos dados é importante conhecer todas as partes integrantes da chamada “*data value chain*” (ou cadeia de valor dos dados, em português), bem como a sua forma mais complexa, o “*big data value chain*” (ou cadeia de valor do *big data*). O conceito de cadeia de valor, em geral, é empregado como uma ferramenta para auxiliar no processo decisório a fim de entender a cadeia de atividades que uma organização realiza para entregar produtos ou serviços valiosos para o mercado (PORTER, 1985).

No caso da cadeia de valor dos dados (*data value chain*), descreve-se o fluxo de informação como uma série de etapas consecutivas e necessárias para geração de valor e *insights* de negócio, sendo cada etapa responsável por determinado processo de transformação sobre os dados. Segundo *framework* de Nguyen e Paczos (2020), esta cadeia é dividida em quatro principais etapas (Figura 11): (i) coleta dos dados, de modo a capturá-los de alguma forma por meio de sensores; (ii) armazenamento ou agregação dos dados, apoiando-se no uso de estruturas de armazenamento que forneçam um acesso fácil e simplificado às informações coletadas; (iii) análise, utilizando-se de ferramentas analíticas a fim de gerar potenciais *insights* de negócio; e, por fim, (iv) uso e monetização, que se trata da aplicação de fato dos *insights* extraídos ao negócio.

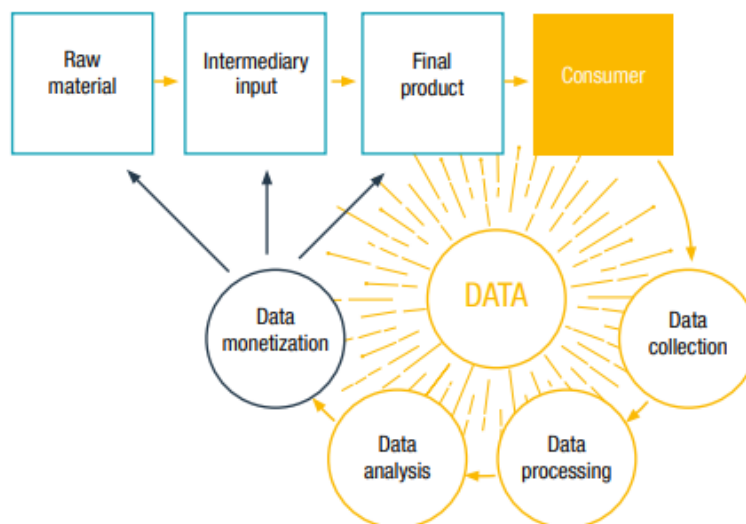
Figura 11 – *Data Value Chain* (Cadeia de valor dos dados)



Fonte: UNCTAD (2019, p. 24).

A dinâmica econômica das plataformas digitais se difere dos modelos de negócio tradicionais que funcionam unicamente sob “*pipeline*” – isto é, o sistema formado por uma cadeia linear de valor que segue um caminho único: fornecedores entregam insumos produtivos, as empresas geram sua produção, colocam o produto ou serviço à disposição do mercado e o cliente compra. O negócio de celulares da Apple é essencialmente um pipeline, por exemplo. Entretanto, quando combinada pela interface e tecnologia de sua App Store, conecta-se aos desenvolvedores de aplicativos e aos usuários do smartphone da empresa, resultando em uma dinâmica de plataforma. Portanto, a Apple atua tanto sob o regime de *pipeline* como por plataforma (VAN ALSTYNE; PARKER; CHOUDARY, 2016). Este novo modelo econômico funciona em uma dinâmica circular, na qual os dados e suas interações em uma rede de conexões funcionam tanto como insumo quanto como fontes de valor (Figura 12).

Figura 12 – Da produção linear ao *feedback loop* na economia digital



Fonte: UNCTAD (2019, p. 39).

Com o desenvolvimento e as mudanças nas práticas de gestão de dados ao longo dos anos, novas tecnologias de armazenamento e processamento de dados vão surgindo, levando a uma nova dinâmica de manipulação de dados, isto é, o *big data*. Alguns autores defendem que com essas mudanças de contexto as organizações são levadas a repensarem a forma que agregam valor aos dados, buscando uma abordagem mais complexa que a cadeia de valor dos dados (*data value chain*) (FAROUKHI *et al.*, 2020). Desta forma, estrutura-se a chamada cadeia de valor do *big data* (*big data value chain*), que possui algumas semelhanças com a *data value chain*, porém com mais etapas que façam sentido com suas características baseadas nos 5 Vs (volume, velocidade, variedade, veracidade e valor). Utilizando o *framework* de Viscont, Larocca e Marconi (2017), as partes são composta por: (i) criação ou captura de dados; (ii) armazenamento de dados; (iii) processamento dos dados, por meio de processos de *data mining* – explicado no tópico 3.1.1 – e análises; (iv) consumo dos dados, por meio de compartilhamento; e, por fim, (iv) monetização dos dados. Conforme Figura 13, cada etapa demonstra um impacto decrescente das características de “volume” e “velocidade” do *big data*, enquanto a característica de “valor” aumenta. Isso ocorre devido ao fato de que nas primeiras etapas da cadeia os dados estão em abundância, porém ainda desestruturados e sem correlações relevantes entre si, o que tem pouco valor em termos de informação útil. Ao longo da cadeia, os dados vão sendo refinados e adquirindo cada vez maior sentido e contexto, o que aumenta a capacidade de utilizar os dados de forma eficiente à sua monetização.

Figura 13 – Big data value chain (Cadeia de valor do big data)

	1. CREATION (Data Capture)	2. STORAGE (Data warehouse)	3. PROCESSING (Data mining & fusion)	4. CONSUMPTION (Visualization & Sharing)	5. MONETIZATION (business plan)
VOLUME	++++	+++	+++	+	+
VELOCITY	+++	+++	+++	++	+
VARIETY	+++	+	+++++	++	+
VERACITY	+	+	++++	++++	++
VALUE	+	+	++	+++	+++++

Fonte: Visconti, Larocca e Marconi (2017, p. 3).

Dentro de cada uma dessas etapas estão atuando empresas e plataformas como Facebook e Google, que oferecem gratuitamente serviços digitais, Airbnb e Uber, intermediando oferta e demanda de determinadas necessidades de serviços, Netflix e Steam, que vendem produtos digitais, Oracle e Microsoft, atuando como empresas desenvolvedoras de sistemas operacionais e ferramentas de manipulação de dados, entre diversas outras. Todas essas funções estão dentro da cadeia de valor do *big data*, porém algumas empresas são capazes de atuar em todas as etapas da cadeia simultaneamente, como por exemplo a Amazon (UNCTAD, 2019).

3.4 PRECIFICAÇÃO DOS DADOS DIGITAIS

A precificação dos dados não se dá de forma simples e homogênea, tendo em vista que suas características também não se dão assim. Na verdade, depende da estrutura de mercado na qual estão inseridos, quem os produz, onde, quando, quão estruturados estão, se são passíveis de combinação com outros dados, entre outras caracterizações. Isso significa, também, que eles são dependentes de contexto para serem entendidos de forma eficiente, não existindo um valor intrínseco isoladamente do ambiente socioeconômico em determinado tempo histórico.

Conforme explica relatório da OCDE (2013a), as informações também são bens de experiência, ou seja, os usuários devem experienciá-las para conseguirem metrificar seu valor. Diferentemente dos bens de consumo comuns, dificilmente existe uma forma de testar a informação e reduzir as incertezas do comprador, o que leva ao paradoxo da informação de Arrow (1962): *“its value for the purchaser is not known until he has the information, but then he has in effect acquired it without cost”*. Portanto, a fim de manter a eficiência de mercado, alguns mecanismos podem ser utilizados para medir a propensão de compra de consumo desses bens, seja através de instituições formais como contratos que garantam uma qualidade suficiente da informação vendida (ARKELOF, 1970) ou instituições informais como reputação de vendedores e resenhas de compradores, que são bastante comuns em lojas de aplicativos ao

disponibilizarem na plataforma um espaço para avaliação do produto ou serviço, por exemplo. Assim busca-se reduzir assimetrias de informação neste mercado (HOLMSTROM, 1985). Outra alternativa seria criar conjuntos de dados de “demonstração”, com apenas uma porção reduzida do total da informação ou uma versão com detalhes limitados, como uma versão anonimizada (MARTENS, 2020).

O fato de o valor dos dados depender do contexto explica por que alguns dados podem ser mais valiosos para determinados atores econômicos e menos para outros. Embora não exista uma única abordagem sobre precificação de dados, sabe-se que a informação a ser extraída dos dados depende do contexto, sendo, portanto, quase impossível determinar seu valor antes de ser aplicado em algo. Portanto, é importante entender também outros fatores que influenciam na capacidade de sua monetização (Figura 14).

Além da disponibilidade de estrutura tecnológica (*hardware*, *software* e redes de conexão) que permita a coleta e armazenamento de informação digital, alguns dos principais fatores que potencializam sua monetização são baseados na capacidade dos dados de serem combinados e recombinados entre si (*data linkage*), a fim de extrair mais *insights* de diferentes conjuntos de informações, bem como a capacidade analítica (*data analytics capacities*) disponível para extrair insumos valiosos dessas informações, o que inclui técnicas e tecnologias específicas. Ou seja, um mesmo conjunto de dados pode gerar diferentes informações dependendo das capacidades analíticas de quem o interpreta (OCDE, 2015). Além disso, o contexto do *big data* é caracterizado por uma superoferta de dados, o que também fortalece o argumento sobre priorização da qualidade e diferenciação dos dados, pois desta forma se agrega mais valor ao bem.

Figura 14 – Fatores que influenciam na determinação do valor dos dados



Fonte: PWC (2019).

Portanto, a monetização dos dados pode surgir de diferentes formas e em contextos diversos, além de existir uma variedade de maneiras de manipular e disponibilizar essas informações dentro do mercado, de acordo com as etapas de tratamento dos dados. Dependendo do modelo de negócio adotado (modelos *data-enhanced* ou *data-enabled*), essas maneiras incluem a venda direta de dados, a melhoria na operacionalização por meio do compartilhamento de dados, o uso de dados para criar e desenvolver novos produtos e serviços, bem como melhorar produtos e serviços pré-existent (NGUYEN; PACZOS, 2020).

3.4.1 Estrutura de mercado e estratégias de preço

Ao compreender o contexto no qual os dados estão inseridos, de fato, no meio econômico, pode-se partir para a próxima construção lógica para entender sua precificação de acordo com determinados objetivos almejados pelas firmas, como, por exemplo lucratividade, posição de mercado, comportamento de preços ao longo de diferentes tipos de produtos e competitividade. Assim como bens físicos, um dos principais fatores que definem os preços dos dados são suas condições de estrutura de mercado (Tabela 15). De acordo com a disponibilidade de poder de mercado para algumas empresas que oferecem determinados produtos e serviços digitais, elas são capazes de determinar preços personalizados para diferentes demandas. No

caso dos monopolistas, por exemplo, possuem poder de mercado o suficiente para aumentar seus lucros sobre o *commodities* de dados e utilizam para tal propósito a estratégia de criar diferentes preços para cada qualidade e nível de detalhe de um conjunto de dados, atendendo níveis distintos de necessidades dos consumidores por discriminação de preços.

Segundo Linde (2009), a economia dos dados está sujeita a mecanismos de mercados imperfeitos devido às suas particularidades características em relação ao seu comportamento econômico – como o fato de ser um bem de experiência capaz de gerar assimetria de informação – e principalmente à sua forma de precificação, portanto os dados estão propensos a gerar ineficiências ou falhas de mercado. A questão da discriminação de preços sobre os dados no mercado será discutida com maiores detalhes no próximo capítulo.

Figura 15 – Características das estruturas de mercado dos dados

Structure	Market share rate	Price determination	Competition	Profit
Monopoly	Some vendors share the market	Have ability to control the price	Weak competition	High
Oligopoly	Very few vendors share the market	Strong ability to control the price	No competition	Very high
Strong competition	Low market share rate	Have no ability to control the price	Malignant competition	Low

Fonte: Liang *et al.* (2018, p. 301).

Além da estrutura de mercado, o próximo passo é entender qual a estratégia de precificação adotada e que definirá o comportamento de maximização de lucro e minimização de custo. De acordo com Muschalle (2012), são definidos seis principais tipos de estratégias de precificação para o mercado de dados:

- **Estratégia de dados gratuitos**, obtidos por meio de dados abertos de autoridades públicas. Embora sua ideia central não seja a troca por retorno monetário, a disponibilidade de dados gratuitamente pode auxiliar na atração de consumidores e estimular o consumo dessas informações, o que em consequência pode atrair também fornecedores de dados comerciais. Além disso, dados gratuitos podem ser integrados com outros dados privados pré-existentes sob posse das firmas (de acordo com sua característica de *data linkage*) e, desta forma, tornarem-se valiosos.
- **Estratégia de preços baseados em uso**, isto é, a métrica do valor está baseada no tempo e na quantidade de dados utilizados. Este exemplo costuma ser aplicado em serviços de internet móvel em telefones, porém, segundo pesquisa de Muschalle (2012), existe uma tendência de desuso desta estratégia devido à perda de persuasão pelo fato do custo marginal neste mercado tender a zero.

- **Estratégia de precificação por pacote** é semelhante à estratégia de preço baseado em uso, porém com um preço fixo. Esta estratégia é estruturada sobre pesquisas em relação a preços baseados em uso (LIANG *et al.*, 2018), a fim de buscar o preço fixo do pacote que maximize os lucros.
- **Estratégia de taxa fixa**, dependendo apenas da variável tempo em sua métrica, além da possibilidade de vender o produto digital apenas uma vez. É comumente utilizada para licenças de *softwares*, como Pacote Office do Windows ou antivírus para computadores.
- **Estratégia de tarifa de duas partes**, sendo uma combinação da estratégia de precificação por pacote e taxa fixa. Neste caso, o consumidor paga uma taxa fixa inicial e valores adicionais de acordo com o que é utilizado ou consumido.
- **Estratégia “freemium”** é uma das novas abordagens utilizadas por grande parte das empresas na economia digital, especialmente as plataformas digitais. Trata-se da ideia de permitir que os usuários usufruam das funções básicas de um serviço ou produto de forma gratuita, mas cobrando por utilidades adicionais. A forma de pagamento desta estratégia pode adotar também qualquer uma das outras estratégias citadas anteriormente. Costuma ser adotada por pequenas empresas de aplicativos em lojas como Apple Store e Play Store (Google) (LIANG *et al.*, 2018).

Observa-se que existe uma gama de diferentes estratégias de preços adotadas pelas empresas no contexto da economia digital, reunindo um conjunto de particularidades em sua dinâmica de funcionamento. Portanto, a análise dos efeitos econômicos do uso dos dados pelas plataformas digitais deve considerar diversos fatores de sua dinâmica, bem como as regulamentações aplicadas sobre cada contexto.

4 AS FALHAS DE MERCADO NO USO DO *BIG DATA* DENTRO DA ECONOMIA DIGITAL

A análise sobre os fenômenos das falhas de mercado presentes na incorporação do *big data* sobre a economia digital vem ganhando bastante notoriedade entre o meio acadêmico nos últimos anos, sendo objeto de estudos recorrentes entre institutos como o National Bureau of Economic Research (NBER) e Harvard Business Review e entre órgãos governamentais de regulamentação de mercado como no Reino Unido, EUA, França, União Europeia, Brasil, entre outros.

Este capítulo buscará trazer a argumentação teórica para a construção de um panorama econômico sobre o qual os fenômenos específicos de falhas de mercado se tornam presentes e de que forma afetam os agentes envolvidos no mercado digital. Adicionalmente, também apresentará algumas das evidências empíricas atuais que ajudam a dar sustentação e robustez para os conceitos a serem trazidos. Por fim, serão observados alguns dados sobre o comportamento dos consumidores e do governo e de que forma têm lidado com esta nova dinâmica da economia digital com possível presença de falhas de mercado.

4.1 EMBASAMENTO TEÓRICO

A construção teórica sobre as falhas de mercado presente no contexto do *big data* sobre a economia digital enfocará em quatro principais fenômenos, interligados entre si: (i) assimetria de informações; (ii) externalidades; (iii) poder de mercado; e (iv) discriminação de preços. Para esta construção lógica, entretanto, serão apresentadas anteriormente as particularidades econômicas resultante do uso dos bens de informação e dos dados digitais dentro da economia.

4.1.1 Características econômicas dos dados

Conforme mostrado no Capítulo 1, os dados vêm ganhando papel estratégico de extrema importância na economia digital, além do vigente processo de sua “commoditização”. Uma vez que a busca pela formação de informação digital utilizando os dados como fonte principal ganha notoriedade entre os atores econômicos, é importante entender como os dados interagem no campo da economia, observando suas particularidades de comportamento e potenciais consequências. Uma vez dito anteriormente por Shapiro e Varian (2000), apesar das mudanças contextuais nas tecnologias da informação, a existência dos fenômenos econômicos mantém

sua dinâmica intacta, bem como alguns conceitos previamente conhecidos no meio acadêmico das ciências econômicas, alterando apenas a perspectiva da análise.

Um dos primeiros desafios do estudo econômico dos dados ou de bens de informação como um todo se dá pelo fato de seu comportamento ser anômalo, pois, para alguns autores no meio acadêmico, acredita-se que “informação viola algumas das características atribuídas a bens de consumo” (BATES, 1990 apud STAHL *et al.*, 2016, p. 10). A própria definição de um bem econômico ainda é tema de grande discussão, entretanto neste trabalho serão considerados os fatores pragmáticos utilizados pela maioria dos autores como a transferibilidade, utilidade e atribuição de valor ou demanda (BIEBERBACH; HERMANN, 1999), requisitos aos quais os bens de informação se adequam.

A primeira característica econômica dos dados é que são considerados não-rivais, isto é, devido às estruturas de armazenamento e transferência de informação digital um mesmo dado pode ser acessado, possuído e utilizado por usuários diferentes ao mesmo tempo. Os bens digitais podem ser replicados e transferidos a custos nulos, o que traz à tona sua característica de não-rivalidade, bem como sua reprodutibilidade a custo marginal zero. Portanto, deve existir um limite no número de vezes que os detentores de dados podem vendê-los até que fiquem sem valor, o que explica parte do preço definido para cobrir os preços de sua produção (STAHL *et al.*, 2016). Aqui o aspecto econômico estratégico mais relevante é mais sobre quem possui o caminho de acesso livre ao invés da propriedade em si. Apesar desta característica e dos possíveis benefícios sociais de uma proliferação de compartilhamentos de dados, os incentivos privados dos atores econômicos são não os compartilhar, a fim de obter vantagens competitivas e proteger questões sensíveis como a privacidade dos indivíduos (CARRIÈRE-SWALLOW; HAKSAR, 2019).

A segunda importante característica dos dados digitais no âmbito econômico também toca em sua não-rivalidade, que é o fato de serem não-excludentes – ou parcialmente não-excludentes. Dado o papel estratégico que os dados adquiriram na economia digital é natural que os agentes tentarão proteger os seus bens de informação a fim de manter vantagens competitivas. Segundo Kitchin (2014), embora um pouco difícil de controlar, ainda é possível limitar o compartilhamento de dados por meio de ferramentas como a aplicação de direitos de propriedade intelectual ou pelos “*paywalls*” – bloqueios digitais que restringem o acesso a um determinado conteúdo através da exigência de pagamento de uma subscrição. Porém, Carrière-Swallow e Haksar (2019) defendem que, na verdade, os dados devem ser considerados parcialmente excludentes, pois para proteger determinados dados é necessário envolver maiores custos para criar mecanismos de segurança de rede e contratar engenharias de estrutura de dados

mais complexas, ou ainda aplicando treinamento para uma mão de obra mais qualificada e criar suas próprias estruturas. Todos esses fatores geram custos, o que nem sempre é acessível a todos os agentes econômicos.

Carrière-Swallow e Haksar (2019) apontam que alguns custos envolvidos na coleta de dados são fixos, como a instalação de uma infraestrutura tecnológica capaz de captar os sinais de informações digitais, enquanto o restante varia de acordo com a quantidade de dados coletado, a depender da dimensão necessária de uma estrutura de armazenamento e o pagamento de qualquer mão de obra envolvida na criação de dados. Ou ainda, Martens (2020) acredita que os custos de coleta são considerados “*sunk costs*” – custos irre recuperáveis – que podem ser amortizados ao longo de diversos usos. Adicionalmente, Farboodi *et al.* (2019) enfatizam a noção de que o custo marginal da coleta de dados é muito baixo, tendo em vista que dados são gerados como subproduto da atividade econômica, como os dados de produção de uma empresa ou de transações de consumidores. Porém, ainda assim o armazenamento e a proteção dos dados são variáveis substanciais de custo para o coletor de dados.

Além dos fatores de custos e a obtenção de vantagem competitiva ao proteger os dados, também é importante considerar a sensibilidade na segurança de informações que envolvem terceiros, como por exemplo dados pessoais de usuários ou consumidores. A criação de uma percepção geral de que as práticas e mecanismos de proteção à privacidade e cibersegurança adotadas pelas firmas e plataformas digitais são insuficientes pode encadear uma externalidade, afetando toda a confiança dos consumidores em compartilharem seus dados e diminuir sua propensão a consumir alguns produtos digitais (CARRIÈRE-SWALLOW; HAKSAR, 2019).

Não é por acaso que os agentes econômicos direcionam grande volume de investimento na segurança e estruturas de armazenamento de seus dados, tendo em vista os potenciais benefícios de sua utilização em ganho de produtividade. Isso, somado ao seu comportamento de não-rivalidade, que permite sua reutilização, leva-nos à próxima importante característica dos dados na economia: capacidade de geração de retornos de escala e escopo. No contexto da economia e das plataformas digitais, o acúmulo de dados pode levar à melhoria na oferta de determinados produtos e serviços, o que conseqüentemente atrai mais usuários, adquirindo cada vez mais dados a serem coletados – isto é, retornos de escala (OCDE, 2015). O uso de ferramentas de *data analytics* como o *Machine Learning (ML)*, conforme citado no tópico 2.1.1, também auxilia nestes ganhos econômicos, pois este tipo de algoritmo é “treinado” e aprende com dados, portanto, quanto maior o volume de informações que o alimenta, mais eficientes e precisos serão os resultados e maiores os benefícios. Para exemplificar este fenômeno, Mayer-Schonberger e Cukier (2013) citam o caso da ferramenta “Google Tradutor”, na qual em 2006

a empresa Google utilizou todas as possíveis traduções existentes na internet da época para treinar seu algoritmo. Com o grande volume de dados que a empresa acumulava, foi capaz de criar traduções muito mais precisas que a de seus concorrentes como da ferramenta “Candide”. Até 2012, o Google Tradutor possuía mais de 60 línguas em seu catálogo.

A própria empresa Google também serve como um ótimo exemplo para explicar outro fenômeno econômico característico no uso dos dados digitais: geração de retornos de escopo. No meio digital e com tantas formas de captação de dados por meio de diferentes fontes (acesso a computadores, celulares, localização via GPS, compras online etc.), as informações captadas podem ser combinadas entre si, gerando mais diversificação de *insights* sobre o mercado. Em muitos casos um conjunto de informação funciona melhor do que de forma isolada, pois contextualiza dados individualizados e lhes agrega mais valor de negócio (OCDE, 2015). No caso da Google, Newman (2013) explica:

It's not just that Google collects data from everyone using its search engine. It also collects data on what they're interested in writing in their Gmail accounts, what they watch on YouTube, where they are located using data from Google Maps, a whole array of other data from use of Google's Android phones, and user information supplied from Google's whole web of online services (NEWMAN, 2013).

Este grande potencial em adquirir vantagens econômicas por meio do uso estratégico dos dados digitais, possibilitado pelo atual estágio de digitalização da economia, também traz consigo uma série de consequências econômicas e desequilíbrios, que serão explicitadas ao longo dos próximos subtópicos neste trabalho. Estes efeitos incluem assimetria de informações, externalidades, poder de mercado e discriminação de preços.

4.1.2 Assimetria de informação

Conforme discutido no tópico sobre valoração e precificação dos dados, tratam-se de um conjunto de bens de informação e experiência, isto é, seu devido valor por parte dos consumidores só pode ser definido após seu consumo e contextualização, o que leva ao paradoxo de Arrow (1962). Essa falta de transparência no processo de valoração desse tipo de bem é definida como “assimetria de informações” no meio econômico, isto é, quando uma das partes em uma transação possui mais informações acerca de um produto ou serviço do que a outra parte, o que leva a falhas de mercado e desequilíbrios alocativos.

Segundo Akerlof (1970), os vendedores que possuem maior conhecimento na qualidade do produto desproporcionalmente em relação ao comprador podem gerar a chamada “seleção

adversa”, uma assimetria de informação *ex ante* à transação que diminui a capacidade do consumidor de avaliar sua “disposição a pagar” e leva sua decisão de compra a uma escolha não-eficiente para si. Na prática, isso pode ocasionar um cenário no qual muitos produtos de baixa qualidade e poucos de alta são vendidos no mercado. No contexto da economia de dados isso é aplicado à incerteza em relação à qualidade do conjunto de dados a ser comprado e o quão útil este poderá ser no contexto que será aplicado para determinado comprador.

O outro caso de assimetria de informação ocorre quando a diferença na proporcionalidade de informações sob posse dos agentes econômicos é manifestada após a transação (*ex post*), sendo que a ação de um dos lados não é verificável. Este fenômeno é denominado “risco moral” e pode estar presente na economia dos dados no lado do consumidor, o gerador de dados (ou *data subject*), devido à falta de clareza sobre as análises que serão feitas sobre seus dados fornecidos ou em relação ao esforço da outra parte sobre a proteção das informações (EL-DARDIRY; DINKOVA; OVERVEST, 2021).

No caso da interação entre usuários individuais e plataformas digitais coletoras de dados é bastante recorrente a existência de assimetrias de informação. Por isso, segundo Bergermann, Bonatti e Gan (2020), a propensão dos usuários em compartilharem suas informações com plataformas depende do nível de detalhes dos dados e qual será sua destinação de uso. Entretanto, pode ocorrer de empresas tornarem obscuras as informações sobre a destinação da troca de dados com seus usuários, levando-os a uma decisão não-eficiente em compartilharem seus dados.

Existe a tese de que o incentivo ao compartilhamento e disponibilização de dados entre todos os atores econômicos pode ser uma alternativa para superar a assimetria de informações e maximizar os benefícios para a sociedade como um todo (OCDE, 2019), entretanto o compartilhamento em massa pode potencializar o risco sobre a privacidade dos indivíduos, o que geraria externalidades.

4.1.3 Externalidades

As externalidades são caracterizadas por ações realizadas por determinados agentes econômicos que possam gerar consequências que afetam outros agentes, impondo custos ou gerando benefícios a eles. Na economia digital e dos dados é bastante comum este tipo de evento, principalmente no que se refere aos denominados “efeitos de rede” e às externalidades negativas sobre a privacidade dos indivíduos.

Os efeitos de rede são externalidades positivas que ocorrem quando o valor de um produto ou serviço aumenta de acordo do número de utilizações por outros indivíduos, o que é bastante comum na implementação e popularização do uso de determinadas tecnologias, como por exemplo a internet. Quanto maior a quantidade de usuários utilizando a ferramenta, maiores as interações entre os agentes econômicos, mais conteúdo e informação são produzidos e, desta forma, as possibilidades de geração de valor através dela também se tornam mais proveitosas. Este tipo de comportamento é bastante comum entre empresas que adotam o modelo de negócio *data-enabled* ou plataformas digitais pois, devido às suas características altamente dependentes de dados para sua operacionalização, acabam usufruindo dos benefícios dos efeitos de rede desta dinâmica. Plataformas de pesquisa na internet, como a Google, utilizam algoritmos que aprendem com as interações de seus usuários, portanto, quanto maior a quantidade de usuários interagindo e “alimentando” a plataforma com dados, melhor será o serviço oferecido.

Os efeitos de economia de escopo e de escala entre as plataformas digitais também são consequências diretas das externalidades de rede, pois com o acúmulo de dados e a melhoria na oferta de seus produtos e serviços, gera-se uma maior atratividade a novos usuários, adquirindo cada vez mais dados a serem coletados, retroalimentando um processo contínuo de melhoria de eficiência. Da mesma forma, efeitos de rede também auxiliam na condução de retornos de escopo na economia digital, pois através desse maior volume de dados também se aumenta a capacidade de combinação entre essas informações, gerando maior diversificação de insumos sobre o comportamento dos consumidores e, conseqüentemente, a criação de novos produtos ou serviços.

Esse grande potencial de escalabilidade e de criação de retornos de escopo e a conseqüente atração de novos usuários também podem ser vistos como potenciais consumidores de novos produtos, o que também atrai mais vendedores e publicidade. Portanto, uma plataforma que atraia um grande público, gera maior exposição de propagandas, o que aumenta sua capacidade de monetização por meio de publicidade via plataforma, gerando novamente uma externalidade positiva para as empresas que funcionam sob estratégias *data-enabled* (EVANS, 2013; NEWMAN, 2015).

Com a ascensão das ferramentas de inteligência artificial e *machine learning*, a capacidade de ampliação dos efeitos de economia de escala e escopo foram aprofundadas, pois o novo ferramental analítico baseado em máquina e algoritmos permite compreender tendências e conclusões sobre o comportamento dos consumidores em uma quantidade de amostragem muito maior do que as capacidades humanas poderiam (GOLDFARB; TREFLER, 2019).

Por outro lado, existem também as externalidades negativas ocasionadas pela dinâmica de funcionamento da economia digital amparada em dados pessoais obtidos de consumidores. Conforme as classificações sobre “*data subject*” apresentadas no tópico 2.1.2, a coleta, compartilhamento e processamento de dados pessoais por determinados agentes econômicos que captam essas informações podem gerar consequências em termos de privacidade para os indivíduos que geraram a informação. Além do lado ético, no viés econômico isso pode significar uma perda de bem-estar do usuário que compartilha seus dados, sem saber o propósito para o qual serão utilizados pela plataforma e o quanto ela preza por estruturas de proteção da privacidade de seus usuários.

Portanto, existem benefícios oriundos do compartilhamento de dados pessoais, como o provisionamento de serviços inovadores e mais personalizados para os interesses de cada indivíduo, além da gratuidade de acesso a plataformas como redes sociais (Facebook, Twitter, Instagram etc.) e ferramentas de pesquisa (Google, Bing, Yahoo, Ask.com etc.). Entretanto, o “custo” invisível está no fornecimento de dados pessoais, que serão analisados, monetizados e talvez disseminados (PFEIFFER, 2019) e, por isso, alguns autores sugerem que o usuário deveria ser pago pela “mão de obra de dados” fornecida, como defendido por Posner e Weyl (2018).

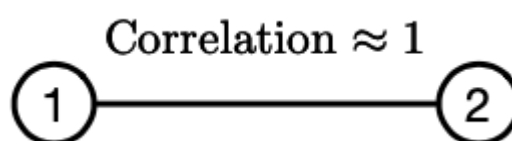
Portanto, quando não há total clareza sobre o uso e a manutenção de dados sensíveis em plataformas existe a possibilidade de agentes privados adotarem decisões econômicas prejudiciais aos *data subjects*, gerando desutilidade sem uma devida compensação de bem-estar. Esse efeito prejudicial aos compartilhadores de informações pessoais pode ocorrer tanto pelo dano à sua privacidade como também pelo uso estratégico dos dados por agentes econômicos com o propósito de extrair valor ou vantagem (CARRIÈRE-SWALLOW; HAKSAR, 2019).

Teoricamente, muitas empresas e plataformas digitais podem se preocupar com a imagem pública que problemas em relação à privacidade de seus usuários podem causar entre os potenciais novos consumidores e a diminuição da propensão em conseguir novos dados para seu funcionamento eficiente, criando, então, incentivos privados para a criação de mecanismos de proteção de dados sensíveis. Entretanto, a valoração da privacidade individual pode ser um tópico difícil de ser definido de forma exata, levando a o que autores denominam de “paradoxo da privacidade”, no qual os indivíduos valorizam sua privacidade na teoria, mas na prática durante suas atitudes cotidianas e no meio digital acabam subestimando seu devido valor (CARRIÈRE-SWALLOW; HAKSAR, 2019).

Neste sentido, existe um outro ponto de preocupação sobre as externalidades em relação à privacidade: quando indivíduos revelam suas informações acabam revelando informação sobre terceiros, mesmo sem consentimento. Essa revelação pode ocorrer tanto de forma direta, por meio de fotos ou conversas que envolvem diferentes indivíduos, ou inferido de forma preditiva por meio de ferramentas analíticas de dados (MACCARTHY, 2010; BERGEMANN; BONATTI; GAN, 2020). Uma vez que uma firma acumula um grande volume de dados sobre consumidores, existe um grande potencial de explorar valor econômico dos usuários, sem a devida compensação de bem-estar.

Acemoglu *et al.* (2019b) trazem essa abordagem de forma mais clara, observando os efeitos das externalidades da privacidade sobre o preço dos dados dos indivíduos. No caso, o “preço” do dado pode ser tanto monetário como um serviço ou produto gratuito, resumido pelo seu “valor de utilidade” na transação. Neste modelo conceitual, os autores utilizam equilíbrio de mercado via jogo de Stackelberg e pode ser resumido da seguinte forma: considerando uma plataforma com dois usuários ($i = 1, 2$) cujos dados estão altamente correlacionados (coeficiente de correlação $\rho \approx 1$) (Figura 16), isto é, após adquirir os dados de um deles a plataforma conseguiria estimar as características do outro usuário. Supondo que a valoração monetária da plataforma sobre os dados “vazados” dos usuários é igual a 1 ($v = 1$), a valoração do primeiro usuário sobre sua privacidade e seus dados “vazados” é $1/2$ ($v = 1/2$) e para o segundo usuário, $v > 0$.

Figura 16 – Valorando a correlação dos dados entre dois usuários



Fonte: Acemoglu (2019a).

Neste cenário, a plataforma oferece preços aos usuários em troca de seus dados. Considerando um cenário sem restrição de mercado ou custos de transação, o usuário 1 sempre irá vender seus dados, pois sua valoração de privacidade é metade do valor da informação pela plataforma, isto é, $\frac{1}{2} < 1$. Tendo em vista que a correlação entre os dados dos dois usuários é alta, a plataforma conseguiria estimar as características do usuário 2 a partir do primeiro, subestimando a valoração do segundo usuário sobre seus dados para um preço praticamente igual a zero. O mesmo fenômeno ocorreria caso a plataforma adquirisse os dados do segundo usuário primeiro. Portanto, no exemplo, a plataforma será capaz de adquirir os dados dos

usuários a um preço muito barato e abaixo do que eles valorizam sobre sua privacidade, sem qualquer compensação, o que gera desequilíbrios distributivos (ACEMOGLU *et al.*, 2019b).

Além disso, um efeito mais sutil desta dinâmica é o fato de que o compartilhamento de dados de um indivíduo altera o valor do dado para plataforma bem como o valor da privacidade de outros indivíduos que tiveram sua informação vazada, reduzindo os incentivos dos usuários a proteger sua privacidade, tendo em vista que suas informações podem ser estimadas através dos dados compartilhados por outros usuários. Uma vez que uma firma acumula um grande volume de dados dos consumidores, o retorno marginal da adição de dados pessoais de um novo usuário fica próximo de zero.

Em busca de entender melhor os efeitos das externalidades da privacidade na economia dos dados atualmente, economistas como Choi, Jeon e Kim (2019) desenvolveram um modelo teórico de privacidade para compreender os efeitos das empresas que coletam dados com o consentimento dos consumidores. Eles concluem que, apesar da existência de ferramentas regulatórias sobre privacidade, as principais utilizadas ainda são ineficientes para tratar os problemas associados à privacidade dos usuários.

4.1.4 Poder de mercado

Apesar dos efeitos aparentemente positivos das externalidades ocasionadas pelos efeitos de rede, esta dinâmica pode fomentar poder e concentração de mercado (*tipping effect*). Conforme apresentado anteriormente na Figura 15, as estruturas de mercado, quando uma empresa possui poder de mercado significa que está em uma posição de relevância neste meio, influenciando ou conduzindo a dinâmica de competitividade e adquirindo a capacidade de estabelecer preços acima do custo marginal (FORGIONI, 2018; TIROLE, 2017). No caso do mercado com alta influência dos dados as plataformas que gozam de uma grande quantidade de usuários – e, conseqüentemente, mais dados – podem usufruir de benefícios que tornam proporcionalmente muito mais difícil para competidores com menores números de usuários se inserirem de forma justa, criando uma situação de poder de mercado (PFEIFFER, 2019).

No relatório “*Unlocking digital competition, Report of the Digital Competition Expert Panel*”, que estuda sobre a competição na economia digital do Reino Unido, Furman *et al.* (2019) justificam porque é tão provável que ocorra concentração de mercado neste contexto. Conforme explicado no tópico anterior, os mercados que envolvem as plataformas digitais estão submetidos a fortes efeitos de economias de escala e escopo, portanto, uma vez que um produto ou serviço digital se torna um sucesso entre consumidores é fácil que a empresa responsável

escale globalmente, além da possibilidade de expandir de forma relativamente mais fácil entre mercados adjacentes. O outro fator se dá pelo fato de os dados atuarem como barreira à entrada de novos concorrentes no mercado, pois firmas que possuem previamente um bom histórico de dados conseguem ter melhores condições de análise sobre os consumidores – tanto aqueles que possuem os dados como aqueles que tiveram seus dados “vazados” de forma preditiva, conforme Acemoglu *et al.*, (2019b). Essa vantagem pode tornar seus produtos cada vez mais eficientes e atrair mais usuários, tirando os incentivos dos consumidores a utilizarem os produtos das firmas menores e levando a uma dinâmica de “*winner takes all*” – quando um competidor se torna hegemônico em um mercado e abarca a grande parcela dos usuários (OCDE, 2012).

Segundo modelo teórico de Kirpalani e Philippon (2020), o compartilhamento de dados por parte dos consumidores gera impactos nos mercados de dois lados cuja interação entre os agentes econômicos ocorre via plataforma digital – o que é o caso aplicado para grandes empresas digitais da atualidade como Amazon ou Google. Os autores concluem que o compartilhamento de dados em excesso por parte dos consumidores leva a cenários de menor competitividade por duas principais razões. O fato citado por Furman *et al.* (2019), da maior capacidade das plataformas compreenderem as características e *preferências* dos consumidores tornarem as plataformas melhores competidoras em comparação com outros tipos de mercados fora da plataforma, tornando cada vez menos atrativo consumir de comerciantes tradicionais (EL-DARDIRY; DINKOVA; OVERVEST, 2021).

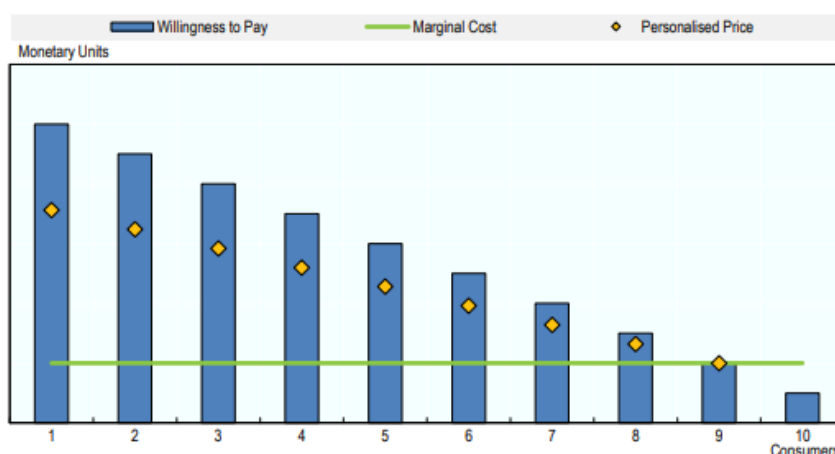
Hagiu e Wright (2020) reforçam a tese de que existem diversas condições na economia digital das quais as plataformas conseguem extrair valor dos dados, levando à cenários de economias de escala. No caso das análises preditivas para entender o comportamento dos consumidores, muitas empresas (como a Google) utilizam algoritmos de inteligência artificial que acumulam uma quantidade massiva de observações e geram grande valor aos dados disponíveis com diversos novos *insights* de negócio.

O uso de algoritmos e inteligência artificial (ou *machine learning*) e o aprofundamento da capacidade de ampliação dos efeitos de economia de escala e escopo vem substituindo parte da mão de obra humana por procedimentos automatizados por máquinas e baseados em dados que, apesar dos custos fixos iniciais altos, possuem custo marginal praticamente nulo. Apesar do consequente alto ganho de produtividade, essa alta capacidade de escalabilidade dos usos de algoritmos também auxilia no aprofundamento do poder de mercado das plataformas maiores (IANSITI; LAKHANI, 2020).

4.1.5 Discriminação de preços

A existência e a atuação de poder de mercado (isto é, mercados imperfeitos) é uma das principais condições para a estratégia de discriminação de preços (MANKIW; TAYLOR, 2012), que ocorre quando uma firma cobra diferentes preços para um mesmo produto ou serviço, dependendo do consumidor que está realizando a transação. O fator que define qual será o preço ofertado a cada consumidor ou grupo de consumidores depende de seu nível de propensão ao consumo, a fim de capturar o máximo possível de excedente de cada classe de consumidor (Figura 17).

Figura 17 – Ilustração da precificação personalizada



Fonte: OCDE (2018, p. 10).

Além da existência de poder de mercado, outra condição necessária para aplicar a discriminação de preços é a capacidade de distinguir as características e o comportamento econômico entre os consumidores. Considerando as transformações tecnológicas que permitiram a digitalização da informação e queda nos custos de geração, coleta e armazenamento de dados, o *big data* criou terreno para a disponibilização de dados massivos sobre o comportamento *online* dos consumidores (sua “pegada digital”). Conseqüentemente, junto aos efeitos de rede, externalidades e poder de mercado, tornou-se mais fácil para firmas e plataformas digitais determinarem de forma mais precisa o comportamento dos consumidores no mercado digital, abrindo espaço para a aplicação de estratégias de “discriminação de preços comportamental” (GOLDFARB; TUCKER, 2019). Com o uso de ferramentas de *data analytics* e técnicas de *data mining* é possível colocar em prática o que se chama de “*profiling*” (conforme explicado no tópico 2.1.1). Quanto maior e mais detalhado o nível de informação sobre os consumidores, maior sua habilidade de aplicar práticas de discriminação de preços – sendo

possível as discriminações de primeiro, segundo e terceiro grau (PIGOU, 1920; TIROLE, 1988; BELLEFLAMME; PEITZ, 2015).

A aplicação, de fato, da discriminação de preços na economia digital ainda é um tema controverso e as evidências empíricas ainda são pouco conclusivas, porém existem alguns casos que ficaram notórios, como o que ocorreu no ano de 2000 no qual um consumidor da Amazon reclamou que recebeu preços diferentes por um DVD no site da plataforma antes e depois de limpar os dados de navegação de seu navegador, o que gerou desconforto entre os clientes. Portanto, a fim de evitar uma reação negativa dos clientes, algumas empresas adotam uma estratégia mais sutil, como oferecer o mesmo preço para todos os consumidores, mas com descontos personalizados para alguns, de acordo com suas características e diferentes propensões a pagar. Outra prática bastante comum é a do “*steering*”, que tem um conceito parecido com a discriminação de preços, porém consiste em estrategicamente dar destaque a produtos específicos de acordo com o grupo de consumidor que está realizando uma pesquisa. Estudos de 2018 da *Competition and Market Authority (CMA)* do Reino Unido e Ipsos *et al.* (2018) encontraram evidências de *steering* entre usuários que recebiam resultados diferentes em algumas páginas da web, de acordo com seu sistema operacional em uso ou sua rota de acesso até a página.

É inegável que, em um ambiente com poder de mercado, as práticas de preços personalizados são benéficas para as firmas e plataformas digitais. Shiller (2014), por exemplo, estima um crescimento entre 0,8% a 12,2% nos lucros da empresa Netflix caso adotassem a estratégia de personalização de preços, utilizando dados dos usuários. Entretanto, considerando os impactos sobre o bem-estar dos consumidores, ainda é incerto se ocorreria um efeito de apropriação – quando os consumidores com maiores propensões a pagar se prejudicam com preços maiores do que um preço padronizado – ou um efeito de expansão de mercado – quando consumidores com menores propensões a pagar acabam sendo beneficiados por preços melhores para seu perfil. Bergmann e Bonatti (2015) demonstram que quando há uma mudança de um preço uniforme para a prática de discriminação de preços, o bem-estar e o excedente do consumidor podem ambos aumentar ou diminuir, bem como um aumentar e o outro diminuir.

4.2 REVISÃO EMPÍRICA

Conforme observado, existe considerável disponibilidade de pesquisas teóricas sobre as falhas de mercado citadas e que dê sustentação para explicar parte da dinâmica distributiva nos mercados digitais baseados em dados, considerando diversos cenários específicos. Entretanto,

é de grande importância observar se essa mesma força acadêmica possui limitações em termos de análises empíricas, a fim de dar maior robustez às ideias defendidas e se os fenômenos são observados na prática. Portanto, serão observados e citados alguns estudos de casos que exemplifiquem cada uma das falhas de mercado citadas no tópico anterior.

4.2.1 Assimetria de informações: caso das ferramentas de pesquisa *online* e do modelo de consentimento informado

Pode-se dizer que a assimetria de informações é uma das bases dos efeitos de falhas de mercado que foram citadas em relação à economia digital e, portanto, algumas sustentações empíricas são necessárias para compreender na prática de que forma os agentes econômicos lidam com o fenômeno. Uma das principais categorias de plataformas digitais que foram observadas empiricamente são as diversas ferramentas de pesquisa disponíveis na internet – similares ao Google, Bing, Yahoo! etc. – que utilizam aspectos de controle e manipulação de informação por meio de um ranqueamento dos resultados de pesquisa de acordo com o perfil do usuário, baseado em seus dados coletados, podendo resultar também em um benefício à própria plataforma. Nas últimas décadas esse tipo de prática vem sendo desenvolvido principalmente em mecanismos de pesquisa de produtos no comércio *online* – como Amazon e eBay –, estruturando também os chamados “sistemas de recomendação” a fim de melhorar a experiência do usuário.

Estudos demonstram que o uso dessas ferramentas pode melhorar a forma de apresentação dos produtos e a experiência do usuário baseado em predições sobre o comportamento e preferência do consumidor (GHOSE; IPEIROTIS; LI, 2012; SANTOS; KOULAYEV, 2017). Segundo Ghose, Ipeirotis e Li (2012a), é possível estimar as preferências dos consumidores de acordo com suas compras realizadas em plataforma *online*, porém outros estudos estimam a possibilidade de determinar as preferências de usuários baseado em outras informações deixadas por sua “pegada digital” em ferramentas de pesquisa, antes mesmo que ocorram as compras de fato (GHOSE; IPEIROTIS; LI, 2012a; SANTOS; KOULAYEV, 2017). Estudos de Chen e Yao (2016), Santos e Koulayev (2017), Koulayev (2014) e Ghose, Ipeirotis e Li, (2012a, 2012b, 2014) utilizam essa abordagem na verificação dos efeitos do ranqueamento em ferramentas de pesquisa sobre as escolhas dos consumidores no setor de hotelaria *online* e encontram uma boa estimativa de margem de efeito de posição.

No estudo de Santos e Koulayev (2017), estima-se um modelo de pesquisa realizado pelo consumidor e propõe-se um método para ranqueamento de hotéis que maximize o

indicador de quantidade de cliques pelos usuários. Bronnenberg, Kim e Mela (2016) estudam um mecanismo de forma semelhante para o mercado *online* de câmeras, definindo o comportamento dos consumidores e as características de sua pesquisa.

A construção de estudos sobre os efeitos dos ranqueamento de resultados de pesquisa necessita de um certo cuidado, pois o ranqueamento é um fator endógeno, isto é, os consumidores prestam mais atenção aos produtos que estão melhor ranqueados, tanto por sua posição como também porque são os mais relevantes para suas preferências (URSU, 2016). Portanto, para identificar o efeito da causalidade dos ranqueamentos, é necessário determinar até onde as escolhas dos usuários dependem da posição do produto ao invés de suas características. Para esta questão, Santos e Koulayey (2016) utilizam ferramentas econométricas de *control function*, método estatístico para corrigir problemas de endogeneidade, enquanto Ghose, Ipeiritis e Li (2014) utilizam modelos de equações simultâneas.

Em geral, o uso das ferramentas de pesquisa ajuda os usuários a encontrar informações específicas que ficam perdidas no meio da quantidade massiva de dados do contexto *big data*, buscando, de certa forma, diminuir parte da assimetria de informação. Entretanto, como demonstrado ao longo das evidências apresentadas, a pesquisa não é “neutra”, existindo a possibilidade de práticas de seleção adversa e risco moral (MARTENS, 2016).

Especificamente no caso das assimetrias a respeito do “risco moral”, que possui uma grande sustentação teórica, o indivíduo gerador de dados possui preocupações sobre quais serão as aplicações de suas informações para quem os coleta. Neste sentido, alguns esforços têm sido buscados para mitigar essas assimetrias, sendo principalmente utilizado o modelo de consentimento informado. Nesta abordagem, os usuários são informados pelos coletores de dados como seus dados serão utilizados antes da transação – como o acesso “gratuito” a uma plataforma – ocorrer de fato. Segundo El-Dardiry, Dinkova e Overvest (2021), esta ideia leva em consideração duas premissas: (a) a privacidade é medida de forma individual; (b) os consumidores são racionais e capazes de tomarem decisões após a informação sobre como serão utilizadas suas informações. Entretanto, o modelo de consentimento do usuário aparenta ser falho e possui uma racionalidade embutida que fica pouco clara. Existe, também, a possibilidade do coletor de dados mentir ou transmitir as informações de consentimento e uso de dados de forma ambígua, colocando em dúvida sua credibilidade (EL-DARDIRY; DINKOVA; OVERVEST, 2021). Em levantamento empírico de Obar e Oeldorf-Hirsch (2018), constata-se que as políticas de privacidade e os termos de serviço são constantemente ignorados pelos usuários. Segundo o levantamento dos pesquisadores, 74% dos indivíduos estudados

pularam a leitura das políticas de privacidade que lhe foram informadas, o que pode ser explicado pelo cálculo de que a leitura completa do texto levaria cerca de 29 a 32 minutos. De fato, utilizando uma análise de regressão, os autores identificaram a sobrecarga de informação como um preditor significativo em relação à propensão do usuário ler os textos. Portanto, segundo as evidências neste caso, o modelo de consentimento informado se demonstra falho para diminuir as assimetrias de informação.

Por outro lado, no contexto dos serviços financeiros, a proliferação de dados pode aliviar o problema de seleção adversa, tornando a oferta de crédito mais eficiente tendo em vista que a assimetria de informações prejudica a população mais vulnerável de países em desenvolvimento (CARRIÈRE-SWALLOW; HAKSAR, 2019). Com o compartilhamento de dados financeiros dos indivíduos entre as instituições do mercado financeiro, evidências empíricas demonstram que, de fato, acaba-se diminuindo os problemas de seleção adversa. Liberman *et al.* (2018) estudam o impacto de uma reforma legal no Chile que impactou o funcionamento dos birôs de crédito, estimando que a medida afetou os tomadores mais vulneráveis, excluindo-os de empréstimos. Outras evidências empíricas como a de Jappelli e Pagano (2002) e Djankov, McLiesh e Shleifer (2007) também fortalecem a noção de que o compartilhamento de dados também resulta em um mercado de crédito mais abrangente. Na economia digital pode ocorrer benefício na coleta de informações *online* sobre os usuários a respeito de seu histórico financeiro digital e auxiliar as instituições de crédito a avaliar sua capacidade de pagamento, apesar de existir a possibilidade de exclusão daqueles que exibem traços de comportamento de risco.

4.2.2 Externalidades: caso dos algoritmos de *machine learning* e os mecanismos de proteção à privacidade online

Os efeitos das externalidades econômicas no contexto da economia digital baseada em dados podem ser explicados, principalmente, por três principais fatores interligados entre si: efeitos de rede (*network effects*), efeitos de economia de escopo e de escala e efeitos sobre a privacidade dos usuários. Apesar da extensa literatura construindo modelos com embasamentos teóricos sobre a dinâmica das externalidades sobre os agentes econômicos na economia de dados digitais, as evidências empíricas ainda se demonstram escassas.

Colin *et al.* (2015) propõe que os efeitos de externalidades de rede na economia digital são explicados pela redução nos custos de transação, tendo em vista que as novas tecnologias digitais fornecem ferramentas rápidas e eficientes para a autenticação das partes em uma

transação e facilitam a construção de conhecimentos sobre a reputação dos agentes. Desta forma, reduz-se as incertezas em relação à desconfiança entre as partes que não se conhecem previamente, conforme estudo empírico de Dyer e Chu (2003). Essa dinâmica se enquadra ao funcionamento de diversas plataformas, como a empresa Uber de transporte pessoal, que interliga de forma segura e eficiente às demandas dos usuários com o serviço de profissionais parceiros, e poderia indiretamente ser utilizado como parte da explicação empírica do argumento sobre os efeitos de rede.

Além dos custos de transação, também se argumenta sobre a influência que o volume de dados sob posse de determinadas firmas pode ter sobre os efeitos de rede e potencialmente levar a falhas de mercado (ABRAHAMSON, 2014; PRÜFER; SCHOTTMÜLLER, 2017). Entretanto, segundo Iansiti (2021), “grande parte da literatura recente sugere que, sobre os efeitos de rede tradicionalmente ditos, o tamanho da rede se trata de apenas um dos fatores que determinam sua força de atuação” (p. 4). Entre esses fatores, incluem-se também a influência das estruturas de rede e a presença de *multi-homing* (técnica na qual é possível se conectar a múltiplos pontos de rede a fim de evitar parada de fluxo caso ocorra falha em uma delas) (ALDRICH; KIM, 2007; CHOI, KIM; LEE, 2010; ZHU *et al.*, 2019; JULLIEN; SANDZANTMAN, 2021). Portanto, Iansiti (2021) acredita que provavelmente a mesma ideia deve se aplicar ao contexto da economia baseada em dados, tendo em vista as escassas sustentações empíricas destes efeitos na realidade.

De fato, os efeitos de rede sobre a economia digital, embora fortemente sustentados por argumentos teóricos, ainda possuem poucas evidências empíricas com as devidas experimentações para explicar sua causalidade e consequências. Um dos poucos materiais com abordagem empírica sobre os efeitos de escala em relação ao aumento no volume de dados obtidos pelas firmas é estruturado por Bajari *et al.* (2019), que analisa o comportamento de ferramentas de *machine learning* da Amazon e sua capacidade de escalabilidade com alimentação de dados. Os autores realizaram um levantamento do impacto da quantidade de dados sobre a performance preditiva do sistema interno da empresa que realiza previsões de vendas no varejo, a fim de validar o denominado “*data feedback loop*”, hipótese que diz que quanto mais dados, melhores serão as previsões da empresa, melhor o atendimento aos consumidores e, conseqüentemente, maior será o fornecimento de dados por usuários (BAJARI *et al.*, 2019). Os resultados do levantamento demonstram que os ganhos são inicialmente significativos para o algoritmo, mas apresentam retornos decrescentes de performance e se esvai rapidamente. Portanto, os ganhos de produtividade dependem mais do tipo de exploração de valor dos dados ao invés de seu volume coletado apenas (BAJARI *et al.*, 2019).

Embora ainda se defenda que não existam evidências suficientes para descartar a existência de poder de mercado gerado pela manipulação estratégica de dados dentro das firmas, as evidências empíricas (ou sua falta) acabam por enfraquecer a hipótese. Assim como no caso das assimetrias de informações, o caso das ferramentas de pesquisa *online* aparenta ser um dos poucos casos com verificação empírica sobre os efeitos de rede e economia de escopo e escala no meio digital. De fato, a lógica de funcionamento das ferramentas de pesquisa (e de suas propagandas direcionadas embutidas) são baseadas na coleta de dados sobre o comportamento de seus usuários, a fim de entender as páginas e produtos mais relevantes para cada termo de pesquisa. Esta forma de alimentar o algoritmo, teoricamente, aumenta a eficiência probabilística nos resultados de pesquisa da ferramenta para melhor atender às necessidades dos usuários, que por sua vez fornecem cada vez mais dados, gerando os efeitos de rede baseados nestes dados (PRÜFER; SCHOTMÜLLER, 2017). Um exemplo para ilustrar este caso é um levantamento realizado por McAfee e Brynjolfsson (2016) que demonstra que um maior número de usuários torna as pesquisas por termos específicos e raros na plataforma Google mais eficientes do que no Bing, que possui um número muito menor de usuários.

Todas essas questões de uso dos dados coletados dos usuários via plataformas gera uma opacidade sobre o funcionamento do mercado digital, o que cria abertura para preocupações e regulamentação em relação à proteção dos direitos dos consumidores e uma busca em metrificar o quanto estão sendo lesados neste processo. Até então se depreende que essa aplicação dos dados como *input* na estratégia de negócios das empresas melhora a eficiência dos produtos e serviços oferecidos, principalmente no sentido de entender as vontades e necessidades dos consumidores do mercado, entretanto o custo para os usuários pode ser a proteção sobre sua privacidade em relação a suas interações dentro do ambiente virtual.

Essa preocupação em relação à preservação da privacidade dos usuários na economia digital e a atenuação das externalidades provocadas por sua dinâmica gerou um movimento por parte do Estado em diversos países de estruturar uma regulamentação especificamente voltada à governança de dados em território nacional. O Comitê Judiciário do Congresso dos Estados Unidos tem demonstrado preocupações sobre os efeitos da coleta de dados, conforme relatório:

[...] ‘the cost of acquiring individual data can be substantially below the value of the information to the platform’. In other words, notwithstanding claims that services such as Google’s Search or Maps products or Facebook are ‘free’ or have immeasurable economic value to consumers, the social data gathered through these services may exceed their economic value to consumers (NADLER; CICILLINE, 2020, p. 46).

Por meio de estudo sobre os efeitos dessas regulamentações sobre a economia digitalizada é possível determinar quais foram as consequências que as restrições tiveram sobre a eficiência de sua dinâmica de operação no mercado.

Trabalhos empíricos mais recentes têm se debruçado sobre os efeitos das regulamentações sobre privacidade de dados, indicando que as externalidades do uso de dados na economia são relevantes na tomada de decisão dos consumidores e das firmas. Estudo empírico de Johnson, Shriver e Du (2020) investigam o impacto da iniciativa de permitir que os usuários norte americanos possam autorizar ou não o uso de *tracking cookies* – arquivos criados por páginas da web que rastreiam a atividade do usuário e conseguem exibir propagandas customizadas de acordo com seus interesses – em seu aparelho. Os autores estimaram uma queda de cerca de US\$ 8,58 por gasto de propaganda em cada consumidor que recusaram os *cookies*.

Outro estudo empírico de Aridor, Che e Salz (2020) demonstra os efeitos da ferramenta de desativação de rastreadores no contexto da regulamentação europeia (GDPR) sobre dados e encontra evidências de que quando alguns usuários optam por desabilitar a coleta de seus dados, torna-se mais fácil para as firmas rastrear e interpretar os dados daqueles consumidores que não desabilitaram os rastreadores (ARIDOR, CHE; SALZ, 2020). Portanto, as ações em relação a compartilhamento de dados por alguns usuários acabam gerando externalidades de informação sobre outros, conforme embasamento teórico defendido por Acemoglu *et al.* (2019b) e Bergemann, Bonatti e Gan (2020). Além disso, os usuários que optam por permitir o uso de rastreadores também podem ser prejudicados em termos de bem-estar, dependendo da forma que os dados coletados serão utilizados. Se os dados são utilizados para atender suas necessidades e vontades em relação a consumo, não haveria grandes problemas, entretanto, a partir do momento que as informações são utilizadas para extração de seu excedente do consumidor (utilizando discriminação de preços, por exemplo), então as externalidades poderiam prejudicá-los (ARIDOR, CHE; SALZ, 2020).

Outro argumento sobre as regulamentações, principalmente a GDPR, é que, de fato, a quantidade de dados coletados dos usuários realmente pode diminuir, o que também gera uma queda no volume de informação fornecida para empresas de propagandas direcionadas *online*. Entretanto, a regulamentação sobre a coleta pode impactar a composição desta indústria, aplicando maiores restrições e riscos nas empresas de pequeno e médio porte do ramo, levando a maior concentração de mercado (GOLDBERG; JOHNSON; SHRIVER, 2019). Portanto, como resultado, este cenário poderia expor os consumidores a outros tipos de malefícios, como a atuação de uma estrutura de mercado monopolística. Além disso, conforme relatório da IHS

Technology (2015), o prejuízo à indústria de propaganda *online* pode também afetar a disponibilização de produtos e serviços gratuitos *online*, tendo em vista que sua grande maioria é sustentada por propagandas digitais.

Neste sentido, com algumas evidências sugerindo que as atuais tentativas de regulamentação da economia digital acabam tendo efeitos adversos e benefícios questionáveis, existe o cenário no qual os efeitos de bem-estar em relação às externalidades de privacidade podem não ser de grande prejuízo aos consumidores considerando o que se ganha em troca. Levantamentos empíricos de Brynjolfsson *et al.* (2019) suportam a noção de que os benefícios sociais oriundos dos novos bens e serviços digitais gratuitos podem superar os malefícios, gerando um maior excedente do consumidor.

Para realizar esta experimentação, os autores elaboram uma nova métrica de PIB (ou GDP, em inglês) que contabiliza os bens digitais novos ou “gratuitos” e que acabam não entrando no sistema tradicional de contas nacionais, denominando-a de GDP-B. Utilizando uma população experimental dos Estados Unidos e o uso da plataforma do Facebook, foi realizado um levantamento no qual os participantes do experimento deveriam escolher entre continuar tendo acesso à rede social ou perder seu acesso durante um mês e ganhar uma recompensa monetária que foi alocada de forma aleatória entre 12 preços. Após a manipulação de um modelo de regressão logística binário, encontraram um preço mediano de propensão a aceitar (*willingness-to-accept* ou *WTA*) se desligar do Facebook por um mês no valor de US\$ 42,17. Utilizando a data de criação do Facebook como ponto de comparação, foi realizado um cálculo estimado das contribuições em termos de bem-estar que a plataforma gerou entre 2003 e 2017, chegando ao valor de US\$ 231 bilhões por ano e uma capacidade de aumentar o GDP-B em termos reais em 1,54%. Portanto, considerando que o Facebook é apenas um dos produtos digitais gratuitos oferecidos atualmente, os impactos podem ser ainda maiores.

Além deste estudo, Goldfarb e Tucker (2012) também trazem evidências sobre os efeitos das regulamentações sobre privacidade e sua capacidade de diminuir o uso dos dados como catalisador de inovações na economia, embora não proponham um nível ótimo de privacidade a ser mantida para favorecer essa dinâmica. Apesar disso, é importante que o enfoque das políticas de privacidade não se pautem apenas na maximização da capacidade de inovação, mas sim na consideração de um equilíbrio com os valores éticos e de privacidade dos indivíduos (GOLDFARB; TUCKER, 2019).

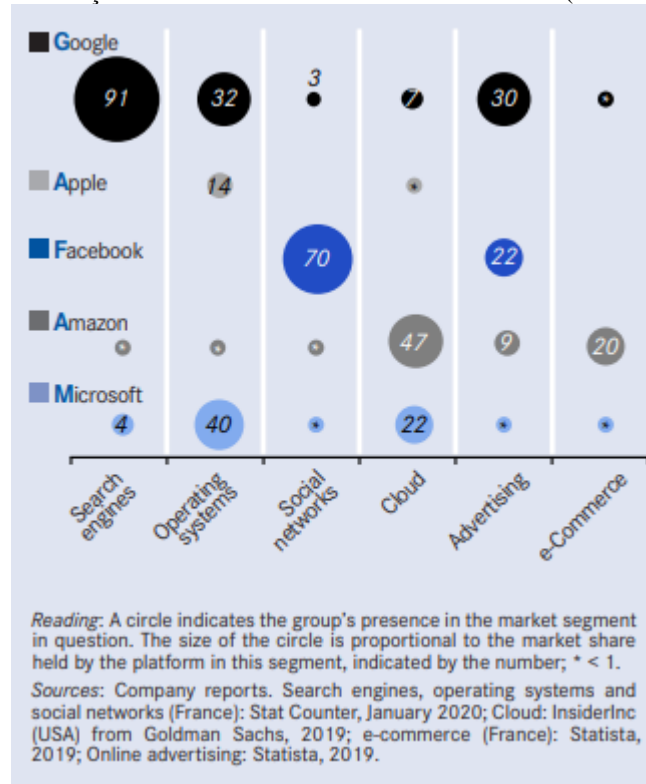
É justamente este apreço ético pela privacidade dos indivíduos e a valorização dos usuários em proteger suas informações pessoais que acaba criando incentivos privados para as firmas terem cuidado em sua governança sobre dados. Diversas evidências empíricas sustentam

a eficácia da reputação *online* em relação ao desempenho de vendas de plataformas *marketplace*, sendo aqueles melhores avaliados os que conseguem estipular maiores preços e, portanto, gerar mais receita (MELNIK; ALM, 2002; LIVINGSTON, 2005; HOUSER; WOODERS, 2005; LUCKING-REILEY *et al.*, 2007). Da mesma forma, estudos empíricos também demonstram que a privacidade está bastante ligada à confiança dos consumidores em relação a quem está coletando suas informações, portanto, quanto maior sua confiança sobre a capacidade do coletor de dados ter uma estrutura eficiente de proteção de dados, maior será sua propensão em compartilhar suas informações e maior o benefício para a plataforma (CAMPBELL; GOLDFARB; TUCKER, 2015).

4.2.3 Poder de mercado: caso das GAFAMs

Um dos argumentos mais preocupantes em relação à economia digital e o desenvolvimento das plataformas digitais é sua grande capacidade de criar estruturas de poder de mercado, tendo em vista as potenciais práticas de aproveitamento das falhas de mercado citadas anteriormente, como a assimetria de informação na manipulação de resultados de pesquisa de forma que favoreça a plataforma ou seus mercados periféricos. As chamadas GAFAMs (Google, Apple, Facebook, Amazon e Microsoft), empresas gigantes da tecnologia que dominam o mercado digital, embora possuam maior *market-share* (participação de mercado) dominante em um setor, geralmente aquele de sua origem, também possuem grande capacidade de expansão para diversos outros mercados (Figura 18).

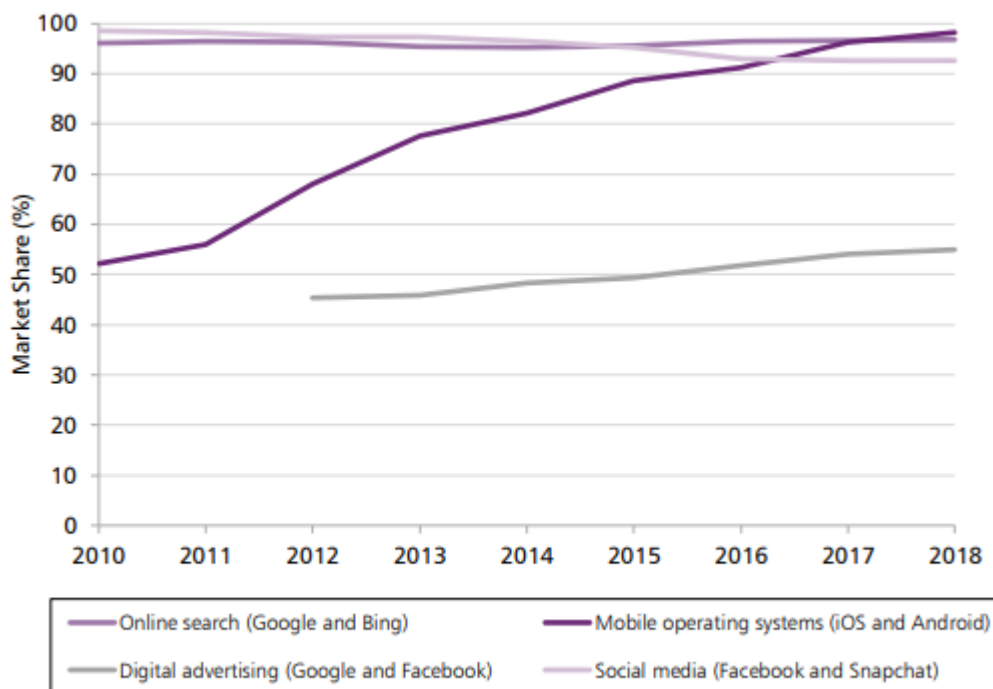
Figura 18 – Posições dominantes e *market share* das GAFM (2019-2020) em %



Fonte: Bourreau e Perrot (2020, p.4).

De fato, segundo Documento do Conselho Administrativo de Defesa Econômica (CADE) e diversos outros relatórios de autoridades do Reino Unido, Comissão Europeia, Austrália e Alemanha, reconhece-se a existências de características capazes de gerar estruturas de poder de mercado entre plataformas digitais no contexto atual da economia digital, criando barreiras de entrada, diminuindo a propensão de ameaça de novos competidores e permitindo a capacidade de entrar em mercados adjacentes (LANCIERI; SAKOWSKI, 2020). Essa dinâmica, que também pode ser observada entre as GAFAMs, é fortemente sustentada pela teoria de que esses mercados são suscetíveis a economias de escala e escopo, o que explicaria a existência de uma ou duas empresas dominantes na maioria dos diversos ramos *online*, como serviços de pesquisa, redes sociais, aplicativos de troca de mensagem etc. (PFEIFFER, 2019). Fuhrman *et al.* (2019), em relatório contendo levantamentos sobre o cenário competitivo do mercado digital do Reino Unido em 2018, trazem alguns dados que reforçam a tese de que o meio digital em seus diferentes ramos costuma ser dominado por poucas empresas (Figura 19), sendo que a pouca competição que existe entre elas está restrita ao mesmo pequeno grupo.

Figura 19 – Indicadores combinados de *market share* das duas atuais empresas líderes em determinados mercados digitais no Reino Unido (2010 - 2018)



Sources: StatCounter,²¹ Comscore,²² and eMarketer and company reports²³

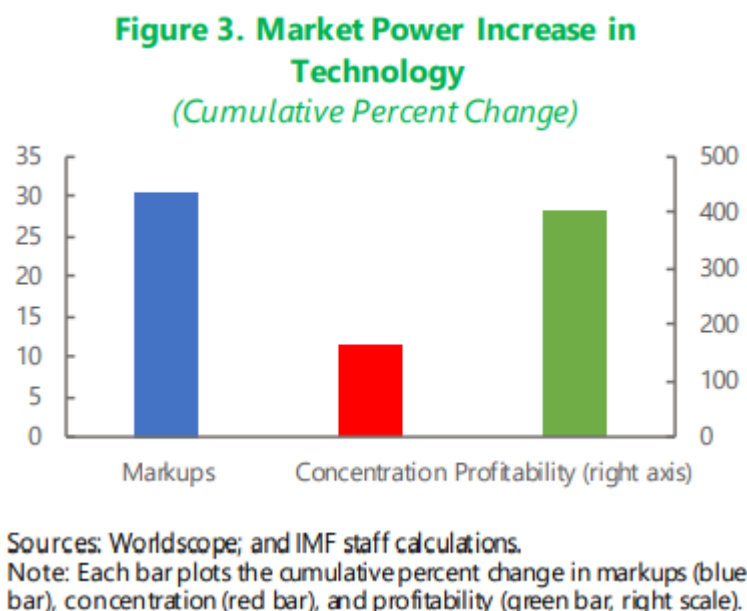
Fonte: Furman *et al.* (2019, p. 25).

Ainda que exista pouca competição, Tirole (2020) cita também o fato de que muitas vezes acabam ocorrendo grandes compras entre si, de forma que a competição é suprimida mais ainda. Um exemplo ocorreu na compra do Instagram pelo Facebook em 2012, duas empresas que estão no mesmo ramo de redes sociais. A Google também realizou movimentos semelhantes quando adquiriu a plataforma Waze, concorrente de seu serviço fornecido via Google Maps. Segundo a UNCTAD (2019), a Google chegou a acumular 230 aquisições ao longo de sua existência, com uma aquisição por semana em determinados períodos. Esse tipo de movimento pode ser visto como um ato de “*killer acquisitions*”, fenômeno empiricamente estudado anteriormente por Cunningham, Ederer e Ma (2020) com evidências dentro da indústria farmacêutica.

A aquisição de empresas, conseqüentemente, leva à posse de um maior volume de dados, o que abre margem para uma maior exploração sobre o mercado digital, tendo em vista que a alocação de valor na economia de dados depende da competição e do poder de mercado (HAKSAR *et al.*, 2021). Em estudo do FMI (2021), demonstra-se concordância com a visão de uma crescente tomada de *market-share* por parte de algumas firmas que detêm poder de mercado no setor de tecnologia (DUVAL, 2021). Segundo o estudo, observando dados da indústria de tecnologia de 82 países entre 1995 e 2016, os *mark-ups* – isto é, o quanto os preços

estabelecidos estão acima do seu custo de produção – aumentaram mais que 30%, bem como a concentração de mercado (mais de 10% globalmente) e a lucratividade (Figura 20). Para ilustrar este cenário, as vendas das quatro maiores empresas contabilizam cerca de 2/3 do total de vendas das 20 principais empresas em 2016 (HAKSAR *et al.*, 2021).

Figura 20 – Aumento do poder de mercado no setor de tecnologia (mudança percentual acumulativa)



Fonte: Duval *et al.* (2021, p. 11).

Apesar dos dados apresentados, o maior desafio sobre a concentração de mercado na economia digital é compreender o quanto este fenômeno afeta o bem-estar dos consumidores e qual é a melhor forma de aplicar uma política regulatória de mitigação dos potenciais efeitos adversos desta dinâmica sem que piore os benefícios aos indivíduos e a eficiência no funcionamento do mercado. Segundo Richter (2021), uma análise descuidada sobre os atos de concentração e suas particularidades econômicas pode levar à aplicação de políticas de competição com consequências que piorem os efeitos negativos, como uma maior concentração econômica, menor inovação, diminuição no desenvolvimento de novos produtos e serviços e desgaste dos efeitos distributivos. Ao mesmo tempo, o funcionamento da economia deve ser instituído sob uma premissa de justiça social, significando que os instrumentos jurídicos devem atuar em favor deste princípio e jamais se omitir em limitar o poder econômico diante das responsabilidades perante a sociedade (FRAZÃO, 2017).

4.2.4 Discriminação de preços: caso da Amazon, Netflix e ZipRecruiter

No contexto da economia digital, a questão relativa à aplicação de práticas de discriminação de preços está intimamente ligada à existência de assimetrias de informação entre as firmas e os consumidores, possuindo um mapeamento de dados sobre as características e preferências de diversos grupos de indivíduos, além do gozo de algum nível de poder de mercado, de forma a permitir ter um maior aproveitamento sobre as diferentes estratégias de preços para um mesmo mercado. Alguns indícios da prática de discriminação de preços aparecem em noticiários, como no caso de reportagem do portal Valor Investe cujo título “Preço de produtos em sites muda quando pesquisado em iPhone, Android e computador” demonstra que sites utilizam dados dos usuários para criar mecanismos de “preços dinâmicos”, realizando um teste para lojas brasileiras como Americanas.com, Submarino, Fast Shop, Casas Bahia e Magazine Luiza e verificando variações de preços (LARGUI, 2019). Outra informação noticiada pela Forbes questionava a possibilidade da empresa Uber utilizar os dados sobre a bateria do telefone do usuário, pois aqueles que estão com menos carga no aparelho tenderão a aceitar preços maiores (CHOWDHRY, 2016).

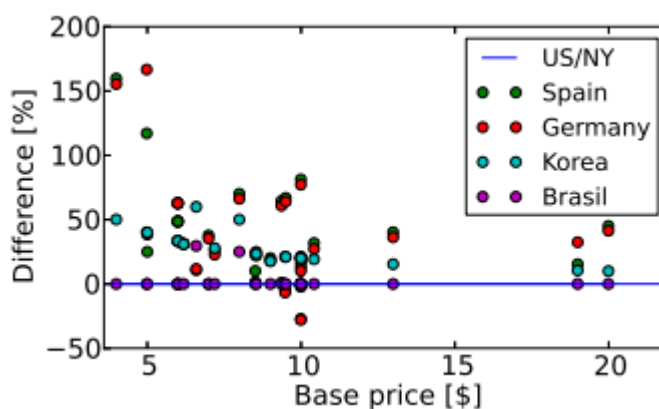
Existe uma aparente existência de condições básicas para a implementação de discriminação de preços no mercado digital e de ser um tema amplamente sustentado pela literatura teórica, entretanto ainda é difícil encontrar evidências robustas e conclusivas sobre a prática de forma generalizada. Segundo relatório do UK Office of Fair Trading (2013) sobre o mercado *online* do Reino Unido, a prática é tecnicamente possível, porém não foram encontradas evidências o suficiente sobre sua aplicação à época. Cinco anos depois, a Competition and Markets Authority (CMA) do Reino Unido replicou o mesmo estudo, de forma atualizada, ainda sem encontrar dados sobre a prática. Outro relatório com dados de *e-commerce* da França buscou evidências de personalização de preços baseada em informações de IP dos consumidores, porém também sem sucesso.

Ainda que as evidências sobre a aplicação de fato de discriminação de preços na economia digital se demonstrem limitadas, existem estudos que observam que existe espaço e ganhos para as firmas caso implementem a prática. Lewis e Rao (2015) realizaram um levantamento sobre a prática de versionamento (discriminação de segundo grau) sobre produtos digitais e encontrou como as empresas podem melhorar seus rendimentos por meio da segmentação de consumidores de alto e baixo valor. Shiller e Waldfoegel (2011) argumentam empiricamente que na indústria da música digital os preços são aplicados de forma uniforme e

acabam ficando abaixo do valor ótimo para maximizar os lucros das firmas, abrindo margem para discriminação de preços.

Apesar do aparente conflito de informações, alguns dados demonstram indícios de discriminação de preços sendo aplicada, mesmo que limitada a um determinado escopo (OCDE, 2018). Mikians *et al.* (2012) demonstram em levantamento empírico a efetividade da discriminação de preços de primeiro grau e da discriminação de pesquisa por meio de ferramentas analíticas de *data profiling*, de acordo com dados sobre transações *online* e localização geográfica dos consumidores, incluindo o Brasil como parte da amostra (Figura 21). No caso do uso da localização geográfica do usuário, os autores reconhecem que as diferenças de preços podem ser atribuídas a outros fatores, como as estruturas logísticas ou questões de propriedade intelectual de cada local.

Figura 21 – Diferenças de preços da Amazon baseado na localização geográfica do consumidor, utilizando os preços de Nova Iorque (EUA) como referência



Fonte: Mikians (2012, p. 83).

Um estudo bastante citado na literatura é o de Shiller (2014), que demonstra como a plataforma Netflix consegue utilizar a personalização de preços (discriminação de terceiro grau) de acordo com algumas informações disponíveis sobre os consumidores e gerar pequenos ganhos incrementais de lucratividade. A precificação baseada no comportamento de navegação *online* do usuário foi capaz de aumentar a lucratividade em 12,2%, enquanto a utilização de dados demográficos gerou um aumento de 0,8%, sendo que ambas estratégias diminuem o agregado do excedente do consumidor, conforme demonstrado na Figura 22. Adicionalmente, considerando um cenário no qual a empresa adota uma estratégia de discriminação mais sutil, oferecendo apenas descontos personalizados para os usuários ao invés de alterar o preço de base, também foi encontrado um aumento sobre a lucratividade, ainda que em menor proporção

(3,19% utilizando os dados de navegação como base e 0,28% com dados demográficos) (Figura 23).

Figura 22 – Simulação das mudanças dos resultados devido à personalização de preços

	Percent Change When Price Based on:	
	Demographics	All Variables
Total Profits	0.79% (0.11)	12.18% (0.68)
Variable Profits	0.14% (0.02)	2.14% (0.12)
Sales (DVDs At-a-Time)	0.85% (1.09)	2.67% (1.23)
Subscribers	0.17% (0.62)	1.65% (0.84)
Aggregate Consumer Surplus	-0.18% (1.12)	-7.75% (1.38)

Bootstrapped standard errors in parentheses. To speed computation simulations in bootstrapping employed a 10 cent increment in price gridsearches.

Fonte: Shiller (2014, p. 31).

Figura 23 – Simulação das mudanças dos resultados devido à personalização de descontos (discriminação de preços de segundo grau otimizada)

	Percent Change When Discounts Based On:	
	Demographics	All Variables
Total Profits	0.28% (0.17)	3.19% (0.51)
Variable Profits	0.05% (0.03)	0.56% (0.09)
Sales (DVDs At-a-Time)	2.56% (0.71)	8.11% (0.89)
Subscribers	2.53% (0.22)	7.61% (0.55)
Aggregate Consumer Surplus	3.38% (0.57)	8.17% (0.64)

Bootstrapped standard errors in parentheses. To speed computation simulations in bootstrapping employed a 10 cent increment in price gridsearches.

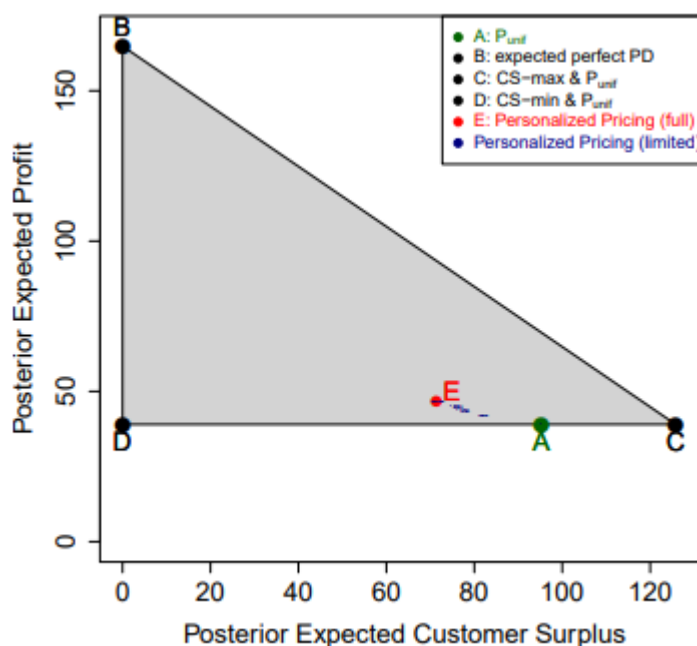
Fonte: Shiller (2014, p. 32).

Outras pesquisas empíricas mais recentes como a de Dubé e Misra (2017) estudam os efeitos sobre de bem-estar no uso de discriminação de preços por meio de ferramentas de *machine learning* em grandes empresas digitais. Em parceria com a ZipRecruiter, plataforma de vagas de emprego, foi realizado um experimento com os dados da empresa e estimou-se que o uso da estratégia de personalização de preços pode gerar, em termos de lucratividade, uma

taxa adicional de 19% em relação à estratégia de precificação uniforme. Por outro lado, no quesito de bem-estar, o uso da discriminação de preços diminuiria o excedente do consumidor em 23% em relação à aplicação do preço uniforme.

Esse movimento é ilustrado pelo “*The Surplus Triangle*” (Figura 24), onde o ponto C representa o máximo de excedente do consumidor, os pontos em azul representam os 62 diferentes esquemas factíveis de segmentações dos consumidores, o ponto E é o cenário de personalização extrema de preços com a segmentação mais granular possível, o ponto A é onde ocorre a precificação uniforme. Tanto em relação ao ponto C como em relação ao ponto A, todos os 62 esquemas de segmentação de preços demonstram queda no excedente do consumidor, chegando a 30% em relação ao ponto A (DUBÉ; MISRA, 2017).

Figura 24 – Triângulo do excedente



Fonte: Dubé e Misra (2017, p. 51).

Apesar dos resultados, o estudo também demonstrou que entre 59,4% e 62,2% dos consumidores entre os 62 diferentes esquemas de segmentação foram beneficiados pelos preços personalizados menores, o que indicaria que uma minoria dos consumidores estaria aumentando os custos da personalização. Portanto, o debate sobre as políticas ligadas à regulamentação sobre o uso de dados e discriminação de preços deveria ser estudada com cuidado (DUBÉ; MISRA, 2017).

4.3 COMO OS ATORES ECONÔMICOS LIDAM COM AS FALHAS DE MERCADO

Enquanto os consumidores vão adentrando a economia digital, utilizando cada vez mais aparelhos móveis, como a popularização dos *smartphones*, e plataformas digitais, como o uso constante de aplicativos de celular, têm-se como resultado a geração e troca massiva de dados. Conseqüentemente, estes dados dos usuários se tornaram cada vez mais atrativos para as empresas, que buscam utilizá-los de forma estratégica para criar modelos de negócio baseados em algoritmos capazes de detectar as preferências e características dos consumidores. Nesta dinâmica, surgem comportamentos que podem gerar conflitos distributivos e éticos, o que implica potenciais malefícios aos indivíduos e, portanto, abrindo margem para a aplicação de jurisdições específicas para regularizar o funcionamento da economia baseada em dados.

Após traçar um panorama geral sobre as diferentes formas que os dados na economia digital podem ser utilizados pelos agentes econômicos (principalmente pelas firmas e a seu benefício) e como cada decisão pode afetar os indivíduos envolvidos nas transações, é importante compreender qual é o cenário atual na presença destes fenômenos e quais são as estratégias e ações dos outros agentes além da perspectiva das firmas para lidar com a nova economia digital. Pela dinâmica simplificada anteriormente, os principais atores econômicos considerados serão: (i) os consumidores, aqueles que alimentam o funcionamento da economia *data-driven* fornecendo seus dados; e (ii) o governo, um agente externo à dinâmica econômica espontânea, mas que tem o papel essencial de regularizar potenciais conflitos de bem-estar na dinâmica entre os consumidores e as firmas.

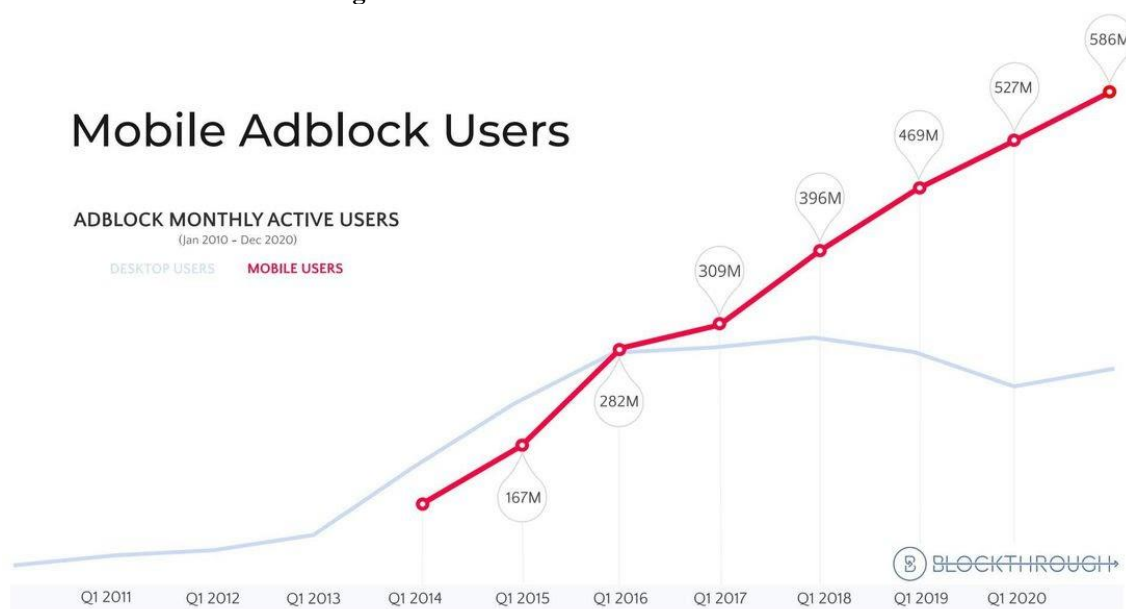
4.3.1 Consumidores

É fato que os novos modelos de negócio digital geram benefícios como flexibilização e diminuição de custos de transação, inovação em termos de produtos e serviços para atender às necessidades dos consumidores, uma grande variedade de bens gratuitos – e sua democratização de acesso –, além da potencialização da importância dada às ferramentas de expressão dos consumidores, seja pela qualidade da transação que está realizando (comentários e avaliações de produtos em *marketplaces*) ou fornecendo *feedback* indireto por meio de interações durante a navegação por plataformas (cliques, tempo de consumo etc). Entretanto, a relação entre os consumidores e o de mercado na economia digital é uma das questões mais importantes no escopo dos fenômenos trabalhados até aqui, pois durante a interação entre empresas ou plataformas digitais e os usuários finais a grande parte da sustentação teórica sobre falhas de

mercado está voltada aos efeitos negativos sobre aqueles que fornecem suas informações em troca de bens digitais, isto é, o consumidor. Isso se dá pelo fato de existirem condições que levam os consumidores a tomarem decisões distorcidas quando considera o *trade off* entre os benefícios de curto prazo e os custos de longo prazo, principalmente no que se refere à questão da privacidade. As evidências demonstram que este comportamento pode ser explicado por dois principais fatores: (a) assimetria de informação; e (b) capacidade cognitiva limitada para processar as informações (ACQUISTI, 2010).

A apresentação incompleta de informação por partes das plataformas que capturam os dados dos usuários acaba gerando um impacto negativo sob sua reputação, aumentando a desconfiança do consumidor e diminuindo sua propensão a compartilhar seus dados, o que seria prejudicial para a firma do ponto de vista econômico. De fato, estudos empíricos vêm demonstrando os efeitos das ferramentas utilizadas pelos usuários para proteger sua privacidade, tais como “*US Do Not Call*” – lista do governo norte-americano que funciona como bloqueador de chamadas telefônicas de telemarketing – ou “*ad-blocks*” – bloqueadores de propagandas na web (GOH; HUI; PNG, 2015; JOHNSON, 2013). Portanto, os consumidores possuem algumas saídas relativamente autônomas para se proteger na economia digital e diminuir os custos de privacidade.

Segundo pesquisa “*2021 PageFair Adblock Report*” da empresa Blockthrough, baseada em usuários dos Estados Unidos, é estimado que 40% dos adultos utilizam ferramentas de *ad-blocks* em seu computador ou smartphone. O uso vem crescendo exponencialmente na internet móvel, chegando a 586 milhões de usuários em 2020, apresentando um crescimento de 80% sobre os usuários *mobile* em comparação com os últimos dois anos (Figura 25). Embora o principal motivo dos usuários utilizarem os *ad-blocks* seja para se proteger de propagandas invasivas (resposta de 81% dos respondentes), 58% também justificam o uso para proteger sua privacidade. Segundo relatório “*Digital 2021: Brazil*” da Data Reportal, o cenário no Brasil não parece estar muito diferente dos usuários norte-americanos. Estima-se que 40,7% dos usuários brasileiros utilizam ferramentas bloqueadoras de anúncios em seu navegador e 50,7% demonstraram preocupação em relação a como as empresas utilizam seus dados pessoais (KEMP, 2021).

Figura 25 – Usuários de *ad-block* em *mobile*

Observa-se que, ao mesmo tempo em que as novas tecnologias de captura de dados podem ser utilizadas para rastrear e analisar uma grande quantidade de informação sobre o mercado de consumo e criar estratégias de direcionamento de produtos ou marketing por parte das firmas, por outro lado também é possível os consumidores utilizarem de tecnologias de informação para se protegerem dessas ações e proteger a privacidade de seus dados. Além dos *ad-blocks*, existem também as chamadas *Privacy Enhancing Technologies (PETs)*. Por meio dessas ferramentas é possível proteger e anonimizar as informações, buscando minimizar os efeitos adversos em relação à privacidade de usuários.

A anonimização dos dados, entretanto, não é uma solução definitiva para o problema de externalidades de privacidade na economia digital. Existem diversas aplicações em que, de fato, dados não identificáveis podem ser úteis, como treinar algoritmos de inteligência artificial de carros automáticos, reconhecimento de imagens, estudos e testes de vacinas e medicamentos, entre outros. Porém, considerando como o dado do consumidor é valorizado na dinâmica da economia *data-driven*, muitas aplicações acabam reduzindo substancialmente o valor da informação caso essa esteja impossibilitada de ser referenciada a um indivíduo. Os custos também são imputados aos consumidores, que devem comprar ou aprender a utilizar ferramentas de anonimização, seja por meio de *PETs* ou deletando *cookies* e navegando em modo anônimo em seu navegador (ACQUISTI, 2010).

Tendo em vista as dificuldades das soluções baseadas em iniciativas e ferramentas dos consumidores, têm-se juntado esforços regulatórios na implementação dos modelos de

consentimento informado, a fim de diminuir as externalidades sobre a privacidade e atacando a assimetria de informação. Entretanto, conforme citado na seção 3.2.1, o estudo empírico de Obar e Oeldorf-Hirsch (2018), conforme previamente citado, demonstrou como as políticas de privacidade e os termos de serviço são constantemente ignorados pelos usuários. Em outro estudo, “2017 Ipsos Globab Trends”, baseado em dados de 23 países, 64% dos respondentes também concordaram que não se importam em ler todos os termos e condições de um site antes de aceitá-los (IPSOS, 2017). No Brasil, a proporção é menor, porém ainda preocupante, chegando em 52% dos respondentes. O número é ainda menor quando se observa quantas pessoas de fato leem os termos e condições, sem considerar uma resposta superestimada dos participantes, chegando a praticamente 1% dos consumidores, o que reforça o “paradoxo da privacidade” (IPSOS, 2014).

4.3.2 Governo

Diante dos riscos citados sobre os potenciais efeitos econômicos e distributivos no comportamento da economia digital, o governo surge como ator principal na busca pelo balanceamento entre os interesses dos agentes econômicos envolvidos nas transações, seja por meio de legislações específicas ou por meio de abertura de investigações de eventos pontuais. Em termos gerais, os principais pontos de atenção em relação à atuação das regulamentações sobre a economia digital *data-driven* pairam questões como o ataque à assimetria de informação existente nas transações digitais, a proteção dos dados pessoais e da privacidade dos consumidores e a defesa da concorrência nestes mercados.

Uma das aplicações de regulamentação na economia digital é a definição de políticas que obriguem as empresas e plataformas digitais a notificarem os usuários de forma clara sobre suas políticas de privacidade, termos de uso e fornecer o máximo possível de informação sobre seu bem ou serviço oferecido, a fim de diminuir a distorção de informação sob posse do consumidor e auxiliar em sua capacidade decisória. Essas informações incluem descrições sobre o negócio, o bem ou serviço, a transação a ser realizada, método de pagamento, como os dados coletados serão tratados, armazenados e utilizados, entre outros. A maior parte dos países integrantes da OCDE, por exemplo, possuem uma legislação que trata sobre a forma em que as empresas devem tratar as informações fornecidas aos consumidores, embora elas se apliquem a qualquer tipo de comercialização, seja online ou offline. Alguns exemplos de nações com boas regras sobre divulgação de informações no mercado digital são Canadá, México, Estados Unidos e União Europeia.

No Canadá existe o “Internet Sales Contract Harmonization Template”, que lista 13 informações necessárias aos consumidores e que devem ser fornecidas por empresas que atuam no mercado online. No México são necessárias informações sobre termos, condições, custos (adicionais ou não), formas de pagamento, bem como dados sobre a empresa (telefone e endereço). No caso dos Estados Unidos, existe a “Federal Trade Commission Act”, que determina algumas informações que as empresas devem divulgar, além de legislações regionais, como na Califórnia, onde há exigências sobre a divulgação de dados sobre os provedores do serviço digital. Na União Europeia as empresas online devem respeitar a “Consumer Rights Directive”, fornecendo informações que preenchem vinte itens estabelecidos, incluindo informações sobre o bem, precificação, direitos e garantias, detalhes de pagamento e contrato, entre outros (OCDE, 2018).

No caso do Brasil, existem mecanismos jurídicos capazes de mitigar as assimetrias de informação no mercado digital, o que foi reforçado com a implementação da recente LGPD. Segundo o Serviço Federal de Processamento de Dados (Serpro), a lei foi criada em com o propósito de criar “um cenário de segurança jurídica, com a padronização de normas e práticas, para promover a proteção, de forma igualitária e dentro do país e no mundo, aos dados pessoais de todo cidadão que esteja no Brasil”, o que inclui o direito ao acesso de informações a respeito de seus dados, incluindo as finalidades pontuais do tratamento, a forma na qual será realizada e sua duração e os dados de contato do controlador, segundo art. 9º, caput e incisos I e II, da LGPD. A fiscalização da aplicação da nova lei e dos direitos sobre dados dos usuários é realizada pela Autoridade Nacional de Proteção de Dados (ANPD), órgão federal criado justamente para aplicação da LGPD, conforme arts. 51, I a XII, 55-B e 55-J, III, IV, XIII, XVI (RICHTER, 2020).

Apesar destes esforços, algumas evidências têm demonstrado que os consumidores pouco valorizam as informações que lhe são fornecidas em relação às políticas de empresas, principalmente no meio digital. Pesquisa de 2019 do Centre for International Governance Innovation (CIGI) em parceria com a Ipsos, “Global Survey on Internet Security and Trust”, realizou um levantamento com mais de 25 mil usuários ao redor do mundo e descobriu que menos da metade da população dos 25 países participantes possuem algum tipo de conhecimento sobre as regras de proteção de seus dados, sendo o Brasil um dos últimos colocados com apenas 33% (CIGI-IPSOS; UNCTAD; INTERNET SOCIETY, 2019). Além disso, mesmo quando as legislações são conhecidas e respeitadas e as informações são fornecidas pelas empresas, poucos usuários leem os textos informativos, portanto, é de grande importância encontrar uma forma eficiente de apresentá-las ao público, buscando o equilíbrio

entre o volume e complexidade capaz de ser absorvido pelos usuários de forma efetiva (OECD, 2018). Alguns autores buscam alternativas, como Malgieri e Custers (2018), que defendem que deveriam existir maiores informações sobre os preços dos dados pessoais que os usuários estão fornecendo nos textos divulgados, a fim de aumentar a atenção dos consumidores para sua leitura. Entretanto, a precificação de dados ainda é extremamente complexa e teria problemas de implementação (MALGIERI; CUSTERS, 2018).

Os esforços em busca de um conjunto de regras que diminuam as assimetrias de informação na economia digital, embora ainda não funcione da forma mais eficiente, também possui o importante papel de estabelecer uma relação de maior confiança com os consumidores. Segundo Furman *et al.* (2019), a construção de um sistema de confiança sobre o uso de dados pode ser um fator propagador de inovação e competição, pois os consumidores se sentiriam mais seguros e propensos a utilizarem mais produtos e serviços digitais, portanto geraria um maior fluxo de dados na economia. Os padrões de valorização sobre a privacidade, portanto, seriam mais um fator de qualidade do bem oferecido diante do mercado.

Entretanto, ainda é necessária uma renovação de valores no mercado digital, saindo da valorização da captura predatória de dados dos consumidores em favor do benefício do econômico e sendo substituída pela imagem positiva que uma boa política de privacidade de uma empresa digital pode causar. Relatório do Reino Unido demonstra que não há evidências de que firmas estejam competindo de forma significativa em termos de proteção à privacidade (FURMAN *et al.*, 2019). Em outro documento do CADE, baseado em relatório do Stigler Center de 2019, “a diminuição de privacidade e o aumento de coleta de dados é uma importante forma de competição não-preço que deve ser examinada por reguladores” (LANCIERI; SAKOWSKI, 2020).

A questão da competitividade, portanto, trata do segundo ponto-chave em relação à atuação do governo em criar políticas mitigatórias das falhas de mercado na economia data-driven. Entretanto, o grande desafio neste cenário é identificar quando ocorre de fato uma situação ou comportamento que se enquadre como anticompetitiva, principalmente quando se trata de um mercado tão particular como o mercado digital. Segundo documento do CADE, após a análise de diversos estudos e relatórios de instituições de diferentes países, a conclusão geral é de que há, sim, problemas de concorrência em mercados digitais e a necessidade das autoridades aplicarem medidas antitruste, utilizando-se a justificativa da defesa do bem-estar do consumidor (*consumer welfare standard*) (LANCIERI; SAKOWSKI, 2020).

Apesar disso, também há de se reconhecer a dificuldade em medir o quão danoso o cenário anticompetitivo ou a intervenção regulatória pode ter sobre o bem-estar dos

consumidores. Em relatório da União Europeia defende-se que as autoridades devem intervir em prol da competitividade em mercados digitais, mesmo sem uma métrica precisa do quanto o consumidor está sendo prejudicado em termos de bem-estar (CRÉMER; DE MONTJOYE; SCHWEITZER, 2019). A mesma ideia é defendida em relatório do Stigler Center, que também propõe uso do bem-estar do consumidor como métrica para políticas de defesa da concorrência, embora também possua o mesmo problema na dificuldade em obter uma forma eficiente de calculá-lo (STIGLER COMMITTEE ON DIGITAL PLATFORMS, 2019).

No caso do Brasil, existe o mesmo cenário em zelar pela justiça social e os benefícios dos consumidores. Segundo Richter (2020), um dos critérios considerados pelas autoridades brasileiras para permitir atos de concentração é que “a parte relevante dos benefícios decorrentes desta operação sejam compartilhados com os consumidores”. Novamente, o grande desafio é a parametrização do bem-estar dos consumidores, além do desafio em parametrizar e identificar os atos de concentração dos agentes de mercado no novo contexto digital. Outras autoridades estrangeiras como a França propõem novas formas de parametrização como número de usuários ativos, número de visitas a sites, número de buscas ou visualizações, tempo ativo online, entre outros fatores característicos do ambiente digital. Portanto, é de grande importância revisar os atuais instrumentos de identificação de atos de concentração (RICHTER, 2020).

Ainda neste sentido, a identificação das particularidades econômicas da economia digital é necessária para compreender os atos anticompetitivos nesta nova dinâmica e, desta forma, criar novos mecanismos reguladores, conforme defendido por relatórios do Reino Unido e Stigler Center. Em relação aos efeitos de rede, associados ao *multi-homing* e *tipping effects*, uma das possíveis saídas é a implementação de estruturas de interoperabilidade entre os setores da economia, isto é, garantir aos usuários o direito de acessar e transferir seus dados entre os concorrentes de mercado, promovendo maior competição e facilitando os custos de troca de provedor (CARRIÈRE-SWALLOW; HAKSAR, 2019). No Brasil, recentemente foi iniciada a implementação do sistema de Open Banking pelo Banco Central do Brasil, que permite o livre compartilhamento de dados dos clientes entre as instituições financeiras, de acordo com sua vontade.

O uso de instrumentos de compartilhamento em massa de dados como um remédio competitivo deve ser analisado com cuidado, pois embora aparentemente tenha efeitos positivos para a autonomia dos consumidores sobre seus dados e aumentar os incentivos concorrenciais entre os atores de mercado, ainda se trata de dados pessoais passíveis de violações aos direitos de privacidade. Nesse sentido, a LGPD e a recente criação da Autoridade Nacional de Proteção

de Dados (ANPD) deve agir de forma cuidadosa e meticulosa para conseguir lidar com as novas adversidades do contexto digital.

5 CONSIDERAÇÕES FINAIS

O trabalho ofereceu uma visão sobre como a evolução das TICs e o avanço da digitalização no mundo conseguiram mudar profundamente a dinâmica de funcionamento das instituições econômicas e a interação entre seus atores, alterando também a forma como a informação e o conhecimento são aplicados para o benefício econômico da sociedade. Apesar disso, os fenômenos tradicionalmente encontrados na literatura microeconômica permanecem e são capazes de explicar alguns dos desequilíbrios que afetam de forma desproporcional os agentes econômicos, o que são resumidos pelas seguintes falhas de mercado: (i) assimetria de informações; (ii) externalidades; (iii) poder de mercado; e (iv) discriminação de preços.

A literatura teórica aborda os fenômenos e seus efeitos distributivos de forma detalhada e possui forte sustentação de documentos oficiais de instituições regulatórias de mercado do governo brasileiro e de nações como Estados Unidos, países da União Europeia, Reino Unido, Japão entre outros. Os dois pontos mais críticos sobre as falhas de mercado se dão pela assimetria de informações na relação entre firma e consumidor, de que forma é extraído o excedente do consumidor e como se dá a perda de bem-estar, gerando implicações éticas sobre a privacidade dos usuários envolvidos nas transações digitais.

Embora os estudos teóricos sejam bastante detalhados, observa-se ainda uma carência de sustentação empírica sobre os fenômenos estudados, tendo em vista a particularidade e a complexidade do funcionamento da economia digital. O valor dos dados na economia *data-driven* é altamente dependente do contexto, portanto as falhas de mercado podem se dar em cenários diferentes e que exigem um estudo minucioso sobre as particularidades de cada um deles, a fim de auxiliar na construção de um *framework* regulatório eficiente e que mitigue de forma adequada os possíveis problemas distributivos entre os agentes econômicos.

Ao mesmo tempo que existe uma limitação da literatura acadêmica em relação a abordagens empíricas dos fenômenos de falhas de mercado que envolvem a economia digital, é interessante também perceber o potencial que o contexto de *big data* também fornece aos estudos experimentais, tendo em vista a grande disponibilidade de amostragens de dados sobre o comportamento da sociedade. Espera-se que, com a implementação de novas estruturas de regulamentação sobre a governança de dados, exista uma observação sobre os resultados destas ações, principalmente no Brasil com a recente LGPD, além da busca por uma maior sustentação empírica robusta sobre de que forma a atuação das GAFAMs e outras plataformas digitais podem gerar implicações distributivas sobre seus usuários. Desta forma, é possível reduzir o grande obscurantismo existente na dinâmica de funcionamento da economia digital em relação

a como os dados dos consumidores são utilizados e como devem ser devidamente recompensados.

REFERÊNCIAS

- AARONSON, Susan Ariel. Data is a development issue. **CIGI Papers**, n. 223, p. 1–32, 2019. Disponível em: <https://www.cigionline.org/publications/data-development-issue/>. Acesso em: 23 set. 2021.
- ABRAHAMSON, Zachary. Essential Data. **Yale Law Journal**, v. 124, n. 3, 2014. Disponível em: <https://digitalcommons.law.yale.edu/ylj/vol124/iss3/7>. Acesso em: 3 nov. 2021.
- ACEMOGLU, Daron et al. **Can we have too much data?** 2019a. Disponível em: <https://voxeu.org/article/can-we-have-too-much-data>. Acesso em: 3 nov. 2021.
- ACEMOGLU, Daron et al. Too Much Data: Prices and Inefficiencies in Data Markets. **Nber Working Paper Series Too**, Cambridge, MA, p. 1689-, 2019b. Disponível em: <https://doi.org/10.3386/w26296>. Acesso em: 28 fev. 2021.
- ACQUISTI, Alessandro. The Economics of Personal Data and Privacy. 2010. In: **OECD Conference Centre**. [s. l.: s. n.], 2010.
- ADLER, Mortmer Jerome. **A Guidebook to Learning: For a Lifelong Pursuit of Wisdom**, 1986.
- AKERLOF, George A. The market for “lemons”: Quality uncertainty and the market mechanism. **Quarterly Journal of Economics**, v. 84, n. 3, p. 488–500, 1970. Disponível em: <https://doi.org/10.2307/1879431>. Acesso em: 2 nov. 2021.
- ALDRICH, Howard E.; KIM, Phillip H. Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search. **Strategic Entrepreneurship Journal**, v. 1, n. 1–2, p. 147–165, 2007. Disponível em: <https://doi.org/10.1002/SEJ.8>. Acesso em: 4 nov. 2021.
- APOLITICAL. **Five ways government is using big data to make your life better**, 2017. Disponível em: <https://apolitical.co/solution-articles/en/five-ways-big-data-changing-government>. Acesso em: 3 nov. 2021.
- ARIDOR, Guy; CHE, Yeon-Koo; SALZ, Tobias. The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR. **SSRN Electronic Journal**, Cambridge, MA, 2020. Disponível em: <https://doi.org/10.2139/ssrn.3522845>. Acesso em: 20 mar. 2021.
- ARROW, Kenneth J. Economic Welfare and the Allocation of Resources for Invention. **The Rate and Direction of Inventive Activity: Economic and Social Factors**, [s. l.], p. 609–626, 1962. Disponível em: <https://econpapers.repec.org/RePEc:nbr:nberch:2144>. Acesso em: 3 nov. 2021.
- BAJARI, Patrick et al. The Impact of Big Data on Firm Performance: An Empirical Investigation. **AEA Papers and Proceedings**, [s. l.], v. 109, p. 1–5, 2019. Disponível em: <https://doi.org/10.1257/pandp.20191000>. Acesso em: 21 out. 2021.

BARBERO, Martina et al. Big data analytics for policy making. **European Commission & Deloitte**, [s. l.], 2016. Disponível em: https://joinup.ec.europa.eu/sites/default/files/dg_digit_study_big_data_analytics_for_policy_making.pdf. Acesso em: 21 set. 2021.

BELLEFLAMME, Paul; PEITZ, Martin. **Industrial Organization: Markets and Strategies**. 2. ed. [S. l.]: Cambridge University Press, 2015. *E-book*.

BERGEMANN, Dirk; BONATTI, Alessandro; GAN, Tan. **The Economics of Social Data***. [S. l.]: arXiv, 2020. Disponível em: <https://doi.org/10.2139/ssrn.3548336>. Acesso em: 28 fev. 2021.

BERGEMANN, Dirk; BONATTI, Alessandro. Selling Cookies. **American Economic Journal: Microeconomics**, v. 7, n. 3, p. 259–294, 2015. Disponível em: <https://doi.org/10.1257/MIC.20140155>. Acesso em: 3 nov. 2021.

BERTOLUCCI, Jeff. **Big Data's New Buzzword: Datafication**. [S. l.], 2013. Disponível em: <https://www.informationweek.com/big-data-analytics/big-data-s-new-buzzword-datafication>. Acesso em: 8 set. 2021.

BIEBERBACH, Florian; HERMANN, Michael. Die Substitution von Dienstleistungen durch Informationsprodukte auf elektronischen Märkten. **Electronic Business Engineering**, p. 67–81, 1999. Disponível em: https://doi.org/10.1007/978-3-642-58663-7_5. Acesso em: 3 nov. 2021.

BOURREAU, Marc; PERROT, Anne. **Digital platforms: Regulate before it's too late**. [S. l.: s. n.], 2020. Disponível em: https://www.cairn-int.info/article-E_NCAE_060_0001--digital-platforms-regulate-before.htm. Acesso em: 25 out. 2021.

BRONNENBERG, Bart J.; KIM, Jun B.; MELA, Carl F. Zooming In on Choice: How Do Consumers Search for Cameras Online?, **Marketing Science**, v. 35, n. 5, p. 693–712, 2016. Disponível em: <https://doi.org/10.1287/MKSC.2016.0977>. Acesso em: 3 nov. 2021.

BRYNJOLFSSON, Erik et al. **GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy**. Cambridge, MA, 2019. Disponível em: <https://doi.org/10.3386/W25695>. Acesso em: 24 out. 2021.

BRYNJOLFSSON, Erik; HITT, Lorin M.; KIM, Heekyung Hellen. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? **International Conference on Information Systems 2011, ICIS 2011**, [s. l.], v. 1, p. 541–558, 2011. Disponível em: <https://doi.org/10.2139/SSRN.1819486>. Acesso em: 27 set. 2021.

BRYNJOLFSSON, Erik; MCAFEE, Andrew. **The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies**. [S. l.]: W. W. Norton & Company, 2016.

CAMPBELL, James; GOLDFARB, Avi; TUCKER, Catherine. Privacy Regulation and Market Structure. **Journal of Economics & Management Strategy**, v. 24, n. 1, p. 47–73, 2015. Disponível em: <https://doi.org/10.1111/JEMS.12079>. Acesso em: 24 out. 2021.

CARRIÈRE-SWALLOW, Yan; HAKSAR, Vikram. **The Economics and Implications of Data: An Integrated Perspective** IMF. [S. l.: s. n.], 2019. Disponível em: <https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2019/09/20/The-Economics-and-Implications-of-Data-An-Integrated-Perspective-48596>. Acesso em: 28 fev. 2021.

CAVANILLAS, José María; CURRY, Edward; WAHLSTER, Wolfgang. **New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe**. [S. l.: s. n.], 2016. Disponível em: <https://doi.org/10.1007/978-3-319-21569-3>. Acesso em: 3 nov. 2021.

CERUZZI, Paul E. **A History of Modern Computing**. Massachusetts: Massachusetts Institute of Technology, 2003. Disponível em: <http://books.google.com/books?hl=en&lr=&id=x1YESXanrgQC&pgis=1>. Acesso em: 24 ago. 2021.

CHEN, Yuxin; YAO, Song. Sequential Search with Refinement: Model and Application with Click-Stream Data. **Management Scienc**, v. 63, n. 12, p. 4345–4365, 2016. Disponível em: <https://doi.org/10.1287/MNSC.2016.2557>. Acesso em: 3 nov. 2021.

CHOI, Hanool; KIM, Sang-Hoon; LEE, Jeho. Role of network structure and network effects in diffusion of innovations. **Industrial Marketing Management**, v. 1, n. 39, p. 170–177, 2010. Disponível em: <https://doi.org/10.1016/J.INDMARMAN.2008.08.006>. Acesso em: 4 nov. 2021.

CHOI, Jay Pil; JEON, Doh Shin; KIM, Byung Cheol. Privacy and personal data collection with information externalities. **Journal of Public Economics**, v. 173, p. 113–124, 2019. Disponível em: <https://doi.org/10.1016/J.JPUBECO.2019.02.001>. Acesso em: 3 nov. 2021.

CHOWDHRY, Amit. Uber: Users Are More Likely To Pay Surge Pricing If Their Phone Battery Is Low. **Forbes**, 2016. Disponível em: <https://www.forbes.com/sites/amitchowdhry/2016/05/25/uber-low-battery/?sh=46f4fcd974b3>. Acesso em: 27 out. 2021.

CIGI-IPSOS; UNCTAD; INTERNET SOCIETY. **CIGI-Ipsos Global Survey on Internet Security and Trust**. [S. l.: s. n.], 2019. Disponível em: <https://www.cigionline.org/cigi-ipsos-global-survey-internet-security-and-trust/>. Acesso em: 30 out. 2021.

CIURIAK, Dan. Unpacking the Valuation of Data in the Data-Driven Economy. **SSRN Electronic Journal**, 2019. Disponível em: <https://doi.org/10.2139/ssrn.3379133>. Acesso em: 10 ago. 2021.

COLIN, Nicolas et al. **The Digital Economy**. [S. l.: s. n.], 2015. Disponível em: <https://www.cae-eco.fr/Economie-numerique>. Acesso em: 24 out. 2021.

CRÉMER, Jacques; DE MONTJOYE, Yves-Alexandre; SCHWEITZER, Heike. **Competition policy for the digital era**. [S. l.: s. n.], 2019. E-book.

- CUNNINGHAM, Colleen; EDERER, Florian; MA, Song. Killer Acquisitions. **SSRN Electronic Journal**, 2020. Disponível em: <https://doi.org/10.2139/SSRN.3241707>. Acesso em: 26 out. 2021.
- DJANKOV, Simeon; MCLIESH, Caralee; SHLEIFER, Andrei. Private Credit in 129 Countries. **Journal of Financial Economics**, v. 12, n. 2, 2007.
- DUBÉ, Jean-Pierre; MISRA, Sanjog. Personalized Pricing and Consumer Welfare. **NBER Working Paper**, Cambridge, MA, 2017. Disponível em: <https://doi.org/10.3386/W23775>. Acesso em: 28 out. 2021.
- DUDA, Richard; HART, Peter; STORK, David. **Pattern Classification**. 2. ed. [S. l.]: Wiley-Interscience, 2000. *E-book*.
- DUVAL, Mr. Romain A. et al. Rising Corporate Market Power: Emerging Policy Issues. **IMF Staff Discussion Notes**, 2021. Disponível em: <https://ideas.repec.org/p/imf/imfstdn/2021-001.html>. Acesso em: 26 out. 2021.
- DYER, Jeffrey H.; CHU, Wujin. The role of trustworthiness in reducing transaction costs and improving performance: Empirical evidence from the United States, Japan, and Korea. **Organization Science**, v. 14, n. 1, p. 57–68, 2003. Disponível em: <https://doi.org/10.1287/ORSC.14.1.57.12806>. Acesso em: 3 nov. 2021.
- EATON, Chris; ZIKOPOULOS, Paul. **Understanding big data: Analytics for Enterprise Class Hadoop and Streaming Data**. [S. l.], 2012. *E-book*.
- EL-DARDIRY, Ramy; DINKOVA, Milena; OVERVEST, Bastiaan. **Policy options for the data economy: a literature review**. [S. l.]: CPB Netherlands Bureau for Economic Policy Analysis, 2021. (CPB background document).
- FARBOODI, Maryam et al. Big Data and Firm Dynamics. **AEA Papers and Proceedings**, [s. l.], v. 109, p. 38–42, 2019. Disponível em: <https://doi.org/10.1257/pandp.20191001>. Acesso em: 3 nov. 2021.
- FAROUKHI, Abou Zakaria et al. Big data monetization throughout Big Data Value Chain: a comprehensive review. **Journal of Big Data 2020 7:1**, [s. l.], v. 7, n. 1, p. 1–22, 2020. Disponível em: <https://doi.org/10.1186/S40537-019-0281-5>. Acesso em: 5 set. 2021.
- FLORIDI, Luciano. **International Encyclopedia of the Social Sciences**. [S. l.: s. n.], 2008. *E-book*.
- FORGIONI, Paula Andrea. **Contratos Empresariais: Teoria Geral e Aplicação**. 3. ed. [S. l.: s. n.], 2018. *E-book*.
- FRAZÃO, Ana. **Direito da concorrência: Pressupostos e Perspectivas**. [S. l.: s. n.], 2017. *E-book*.
- FURMAN, Jason et al. **Unlocking digital competition: Report of the Digital Competition Expert Panel** HM Treasury. Londres: [s. n.], 2019. Disponível em:

<https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel>. Acesso em: 2 nov. 2021.

GHOSE, Anindya; IPEIROTIS, Panagiotis G.; LI, Beibei. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. <https://dx.doi.org/10.1287/mksc.1110.0700>, [s. l.], v. 31, n. 3, p. 493–520, 2012a. Disponível em: <https://doi.org/10.1287/MKSC.1110.0700>. Acesso em: 3 nov. 2021.

GHOSE, Anindya; IPEIROTIS, Panagiotis G.; LI, Beibei. Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue. <https://dx.doi.org/10.1287/mnsc.2013.1828>, [s. l.], v. 60, n. 7, p. 1632–1654, 2014. Disponível em: <https://doi.org/10.1287/MNSC.2013.1828>. Acesso em: 3 nov. 2021.

GHOSE, Anindya; IPEIROTIS, Panagiotis G.; LI, Beibei. **Surviving Social Media Overload: Predicting Consumer Footprints on Product Search Engines**. [S. l.: s. n.], 2012b. Disponível em: https://www.researchgate.net/publication/260298532_Surviving_Social_Media_Overload_Predicting_Consumer_Footprints_on_Product_Search_Engines. Acesso em: 3 nov. 2021.

GOH, Khim Yong; HUI, Kai Lung; PNG, Ivan P.L. Privacy and marketing externalities: Evidence from Do Not Call. **Management Science**, v. 61, n. 12, p. 2982–3000, 2015. Disponível em: <https://doi.org/10.1287/MNSC.2014.2051>. Acesso em: 3 nov. 2021.

GOLDFARB, Avi; TREFLER, Daniel. Artificial intelligence and international trade. **The economics of artificial intelligence: an agenda**. Chicago: University of Chicago Press, 2019.

GOLDFARB, Avi; TUCKER, Catherine. **Digital economics**. Cambridge, MA: [s. n.], 2019. Disponível em: <https://doi.org/10.1257/jel.20171452>. Acesso em: 28 mar. 2021.

GOLDFARB, Avi; TUCKER, Catherine. Shifts in Privacy Concerns. **American Economic Review**, v. 102, n. 3, p. 349–353, 2012. Disponível em: <https://doi.org/10.1257/AER.102.3.349>. Acesso em: 3 nov. 2021.

GOLDBERG, Samuel; JOHNSON, Garrett; SHRIVER, Scott. Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. **SSRN Electronic Journal**, [s. l.], 2019. Disponível em: <https://doi.org/10.2139/SSRN.3421731>. Acesso em: 3 nov. 2021.

GRAHAM, Stephen D.N. Software-sorted geographies. **Progress in Human Geography**, v. 29, n. 5, p. 562–580, 2005. Disponível em: <https://doi.org/10.1191/0309132505PH568OA>. Acesso em: 3 nov. 2021.

HAGIU, Andrei; WRIGHT, Julian. When Data Creates Competitive Advantage. **Harvard Business Review**, v. 98, n. 1, p. 94–101, 2020. Disponível em: <https://hbr.org/2020/01/when-data-creates-competitive-advantage>. Acesso em: 16 out. 2021. Acesso em: 3 nov. 2021.

HAKSAR, Vikram et al. Towards a Global Approach to Data in the Digital Age. **IMF Staff Discussion Notes**, v. 2021, n. 5, 2021. Disponível em: <https://doi.org/10.5089/9781513599427.006.A001>. Acesso em: 19 out. 2021.

HART, Jeffrey A; REED, Robert R; BAR, François. The building of the internet. Implications for the future of broadband networks. **Telecommunications Policy**, v. 16, n. 8, p. 666–689, 1992. Disponível em: [https://doi.org/10.1016/0308-5961\(92\)90061-S](https://doi.org/10.1016/0308-5961(92)90061-S). Acesso em: 25 ago. 2021.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. [S. l.: s. n.], 2011. *E-book*.

HEY, Jonathan. **The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link**. [S. l.], 2004. Disponível em: <http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf>. Acesso em: 8 set. 2021.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. Jim Gray on eScience: A Transformed Scientific Method. **The Fourth Paradigm: Data-Intensive Scientific Discovery** - Microsoft Research, [s. l.], 2009. Disponível em: <http://sites.nationalacademies.org/NRC/index.htm;ComputerScienceandTelecommunicationsBoard,http://sites.nationalacademies.org/cstb/index.htm.-> http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt. Acesso em: 3 nov. 2021.

HILBERT, Martin. Quantifying the Data Deluge and the Data Drought. **SSRN Electronic Journal**, [s. l.], 2015. Disponível em: <https://doi.org/10.2139/SSRN.2984851>. Acesso em: 3 nov. 2021.

HOLMSTROM, Bengt. **The provision of services in a market economy**. Cambridge: Cambridge University Press, 1985.

HOUSER, Daniel; WOODERS, John. Reputation in Auctions: Theory, and Evidence from eBay. **Journal of Economics & Management Strategy**, v. 15, n. 2, p. 353–369, 2005. Disponível em: <https://doi.org/10.1111/J.1530-9134.2006.00103.X>. Acesso em: 4 nov. 2021.

HOW DATA WILL TRANSFORM BUSINESS. Direção: Philip Evans. [S. l.: s. n.], 2013. Disponível em: https://www.ted.com/talks/philip_evans_how_data_will_transform_business. Acesso em: 3 nov. 2021.

IANSITI, Marco; LAKHANI, Karim. **Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World**. Boston: Harvard Business Review Press, 2020. *E-book*.

IANSITI, Marco. The Value of Data and Its Impact on Competition. **SSRN Electronic Journal**, v. 22, n. 2, 2021. Disponível em: <https://doi.org/10.2139/SSRN.3890387>. Acesso em: 24 out. 2021.

IHS TECHNOLOGY. **Paving the way: how online advertising enables the digital economy of the future**. [S. l.: s. n.], 2015.

IPSOS. **Ipsos Global Trends 2017**. [S. l.: s. n.], 2017. Disponível em: <https://www.ipsos.com/en-th/ipsos-global-trends-2017>. Acesso em: 4 nov. 2021.

JAMES, Gareth et al. **An Introduction to Statistical Learning: with Applications in R**. [S. l.]: Springer, 2013. *E-book*.

JAPPELLI, Tullio; PAGANO, Marco. Information sharing, lending and defaults: Cross-country evidence. **Journal of Banking & Finance**, v. 26, n. 10, p. 2017–2045, 2002. Disponível em: <https://econpapers.repec.org/RePEc:eee:jbfina:v:26:y:2002:i:10:p:2017-2045>. Acesso em: 3 nov. 2021.

JIN, Xiaolong et al. Significance and Challenges of Big Data Research. **Big Data Research**, v. 2, n. 2, p. 59–64, 2015. Disponível em: <https://doi.org/10.1016/j.bdr.2015.01.006>. Acesso em: 6 set. 2021.

JOHNSON, Garrett; SHRIVER, Scott; DU, Shaoyin. Consumer Privacy Choice in Online Advertising: Who Optes Out and at What Cost to Industry? **SSRN Electronic Journal**, 2020. Disponível em: <https://doi.org/10.2139/SSRN.3020503>. Acesso em: 4 nov. 2021.

JOHNSON, Justin P. Targeted advertising and advertising avoidance. **The RAND Journal of Economics**, v. 44, n. 1, p. 128–144, 2013. Disponível em: <https://doi.org/10.1111/1756-2171.12014>. Acesso em: 31 out. 2021.

JULLIEN, Bruno; SAND-ZANTMAN, Wilfried. The Economics of Platforms: A Theory Guide for Competition Policy. **Information Economics and Policy**, v. 54, 2021. Disponível em: <https://doi.org/10.1016/J.INFOECOPOL.2020.100880>. Acesso em: 3 nov. 2021.

KEMP, Simon. **DIGITAL 2021: BRAZIL**. [S. l.], 2021. Disponível em: <https://datareportal.com/reports/digital-2021-brazil>. Acesso em: 1 nov. 2021

KENNEDY, Joseph. **The data-driven economy**. US Chamber of Commerce Foundation. [S. l.: s. n.], 2014.

KIRPALANI, Rishabh; PHILIPPON, Thomas. Data Sharing and Market Power with Two-Sided Platforms. **NBER Working Paper 28023**, [s. l.], 2020. Disponível em: <https://www.wsj.com/articles/amazon-tech-startup-echo-bezos-alex-a-investment-fund-11595520249>. Acesso em: 3 nov. 2021.

KITCHIN, Rob; DODGE, Martin. **Code/Space: Software and Everyday Life**. Massachusetts: The MIT Press, 2011.

KITCHIN, Rob. **Cyberspace: The world in the wires**. [s. l.], 1998. Disponível em: <http://www.lavoisier.fr/notice/frRWORXSLA6RWLKO.html>. Acesso em: 25 ago. 2021.

KITCHIN, Rob. **The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences**. Londres: SAGE Publications, 2014.

KNOX, Paul L.; CASTELLS, M. The Informational City. *Information Technology, Economic Restructuring, and the Urban-Regional Process*, v. 161, n. 1, p. 94, 1988. Disponível em: <https://doi.org/10.2307/3059942>.

KOULAYEV, Sergei. Search for differentiated products: identification and estimation. **The RAND Journal of Economics**, v. 45, n. 3, p. 553–575, 2014. Disponível em: <https://doi.org/10.1111/1756-2171.12062>. Acesso em: 3 nov. 2021.

KPMG. Experiência do consumidor: Cresce a preocupação com proteção de dados. **KPMG Brasil**, 2021. Disponível em: <https://home.kpmg/br/pt/home/insights/2021/03/experiencia-do-consumidor-tendencias.html>. Acesso em: 15 nov. 2021.

LANCIERI, Filippo Maria; SAKOWSKI, Patrícia Alessandra. **Concorrência em mercados digitais: uma revisão dos relatórios especializados**. [S. l.: s. n.], 2020. Disponível em: <https://cdn.cade.gov.br/Portal/centrais-de-conteudo/publicacoes/estudos-economicos/documentos-de-trabalho/2020/documento-de-trabalho-n05-2020-concorrencia-em-mercados-digitais-uma-revisao-dos-relatorios-especializados.pdf>. Acesso em: 24 out. 2021.

LANEY, Doug. 3D data management: Controlling data volume, velocity and variety. **META Group Research Note**, v. 6, n. 70, 2001. Disponível em: <https://www.bibsonomy.org/bibtex/742811cb00b303261f79a98e9b80bf49>. Acesso em: 20 set. 2021.

LAPOWSKY, Issie. **How Cambridge Analytica Sparked the Great Privacy Awakening**. [S. l.], 2019. Disponível em: <https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening/>. Acesso em: 26 set. 2021.

LARGUI, Nathália. Preço de produtos em sites muda quando pesquisado em Iphone, Android e computador. **Valor Investe**, 2019. Disponível em: <https://valorinveste.globo.com/objetivo/gastar-bem/noticia/2019/10/07/preco-de-produtos-em-sites-muda-quando-pesquisado-em-iphone-android-e-computador.ghtml>. Acesso em: 27 out. 2021.

LEVIN, Jonathan. **The economics of internet markets**. In: Advances in economics and econometrics: tenth world congress, volume I: economic theory. [S. l.: s. n.], 2011. p. 48–75. Disponível em: <https://doi.org/10.1017/CBO9781139060011.003>. Acesso em: 24 ago. 2021.

LEWIS, Randall A.; RAO, Justin M. The unfavorable economics of measuring the returns to advertising. **Quarterly Journal of Economics**, v. 130, n. 4, p. 1941–1973, 2015. Disponível em: <https://doi.org/10.1093/QJE/QJV023>. Acesso em: 3 nov. 2021.

LIANG, Fan et al. A Survey on Big Data Market: Pricing, Trading and Protection. **IEEE Access**, v. 6, p. 15132–15154, 2018. Disponível em: <https://doi.org/10.1109/ACCESS.2018.2806881>. Acesso em: 3 nov. 2021.

LIBERMAN, Andres et al. **The Equilibrium Effects of Information Deletion**: Evidence from Consumer Credit Markets. Cambridge, MA, 2018. Disponível em: <https://doi.org/10.3386/W25097>. Acesso em: 3 nov. 2021.

LINDE, Frank. Ökonomische Besonderheiten von Informationsgütern. **Wissens- und Informationsmanagement**, p. 291–320, 2009. Disponível em: https://doi.org/10.1007/978-3-8349-6509-7_14. Acesso em: 3 nov. 2021.

LIVINGSTON, Jeffrey A. How valuable is a good reputation? A sample selection model of Internet auctions. **Review of Economics and Statistics**, v. 87, n. 3, p. 453–465, 2005. Disponível em: <https://doi.org/10.1162/0034653054638391>. Acesso em: 3 nov. 2021.

LOUKIDES, Mike. **What is data science?** . [S. l.], 2010. Disponível em: <https://www.oreilly.com/radar/what-is-data-science/>. Acesso em: 3 nov. 2021.

LUCKING-REILEY, David et al. Pennies from eBay: The determinants of price in online auctions. **Journal of Industrial Economics**, v. 55, n. 2, p. 223–233, 2007. Disponível em: <https://doi.org/10.1111/J.1467-6451.2007.00309>. Acesso em: 3 nov. 2021.

MACCARTHY, Mark. New Directions in Privacy: Disclosure, Unfairness and Externalities. **ISJLP**, v. 6, p. 425, 2011.

MALGIERI, Gianclaudio; CUSTERS, Bart. Pricing privacy – the right to know the value of your personal data. **Computer Law and Security Review**, [s. l.], v. 34, n. 2, p. 289–303, 2018. Disponível em: <https://doi.org/10.1016/j.clsr.2017.08.006>. Acesso em: 10 ago. 2021.

MANKIW, Gregory; TAYLOR, Mark P. **Economics**. 5. ed. 2012. *E-book*.

MANYIKA, James et al. **Big data: The next frontier for innovation, competition, and productivity**. 2011. Disponível em: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>. Acesso em: 3 nov. 2021.

MARTENS, Bertin. **An Economic Policy Perspective on Online Platforms**. 2016. Disponível em: <https://ec.europa.eu/jrc/sites/default/files/JRC101501.pdf>. Acesso em: 19 out. 2021.

MARTENS, Bertin. Data Access, Consumer Interests and Social Welfare: An Economic Perspective. **SSRN Electronic Journal**, 2020. Disponível em: <https://doi.org/10.2139/ssrn.3605383>. Acesso em: 13 out. 2021.

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. **Big Data: A Revolution That Will Transform How We Live, Work, and Think**. 2013. *E-book*.

MCKINSEY. **Fueling growth through data monetization**, 2018. Disponível em: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/fueling-growth-through-data-monetization>. Acesso em: 30 set. 2021.

MELNIK, Mikhail I; ALM, James. Does a Seller's eCommerce Reputation Matter? Evidence from eBay Auctions. **The Journal of Industrial Economics**, v. 50, n. 3, p. 337–349, 2002. Disponível em: <https://doi.org/10.1111/1467-6451.00180>. Acesso em: 4 nov. 2021.

MÉNIÈRE, Yann et al. **Patents and the Fourth Industrial Revolution (EPO) European Patent Office**. [S. l.: s. n.], 2020. Disponível em: <https://www.lemoci.com/wp-content/uploads/2017/12/Patents-and-the-Fourth-industrial-Revolution-2017.pdf>. Acesso em: 27 set. 2021.

MERELLI, Emanuela; RASETTI, Mario. Non locality, topology, formal languages: New global tools to handle large data sets. **Procedia Computer Science**, v. 18, p. 90–99, 2013. Disponível em: <https://doi.org/10.1016/J.PROCS.2013.05.172>. Acesso em: 3 nov. 2021.

MIKIANS, Jakub et al. Detecting price and search discrimination on the Internet. In: 11th ACM Workshop on hot topics in networks (HOTNETS-XI). **Anais eletrônicos**. p. 79–84, 2012. Disponível em: <https://doi.org/10.1145/2390231.2390245>. Acesso em: 27 out. 2021.

MPDFT. **MPDFT recomenda providências à Netshoes após vazamento de quase 2 milhões de dados de clientes**, 2018. Disponível em: <https://www.mpdft.mp.br/portal/index.php/comunicacao-menu/noticias/noticias-2018/9775-mpdft-recomenda-providencias-a-netshoes-apos-vazamento-de-quase-2-milhoes-de-dados-de-clientes>. Acesso em: 3 nov. 2021.

MUSCHALLE, Alexander. Pricing Approaches for Data Markets. **Lecture Notes in Business Information Processing**, v. 154, p. 129–144, 2012. Disponível em: https://doi.org/10.1007/978-3-642-39872-8_10. Acesso em: 19 set. 2021.

NADLER, Jerrold; CICILLINE, David N. Investigation of Competition in Digital Markets. **House Committee on the Judiciary**, [s. l.], p. 1-450, 2020. Disponível em: <https://www.cliffordchance.com/briefings/2020/10/US-House-Report-on-Competition-in-Digital-Markets-Focuses.html>. Acesso em: 4 nov. 2021.

NADLER, Jerrold; CICILLINE, David N. Investigation of Competition in Digital Markets. **House Committee on the Judiciary**, [s. l.], p. 1–450, 2020. Disponível em: <https://www.cliffordchance.com/briefings/2020/10/US-House-Report-on-Competition-in-Digital-Markets-Focuses.html>. Acesso em: 4 nov. 2021.

NEWMAN, Abraham L. What the “right to be forgotten” means for privacy in a digital age. **Science**, v. 347, n. 6221, p. 507–508, 2015. Disponível em: <https://doi.org/10.1126/SCIENCE.AAA4603>. Acesso em: 3 nov. 2021.

NEWMAN, Nathan. Taking on Google’s Monopoly Means Regulating Its Control of User Data. **HuffPost**, 2013. Disponível em: https://www.huffpost.com/entry/taking-on-googles-monopol_b_3980799. Acesso em: 17 set. 2021.

NGUYEN, David; PACZOS, Marta. Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective. **OECD Digital Economy Papers**, v. 297, 2020. Disponível em: <http://www.oecd.org/termsandconditions>. Acesso em: 26 mar. 2021.

NOVIKOV, Sergey V. Data science and big data technologies role in the digital economy. **TEM Journal**, v. 9, n. 2, p. 756–762, 2020. Disponível em: <https://doi.org/10.18421/TEM92-44>. Acesso em: 10 ago. 2021.

O’NEILL, Judy E. Role of ARPA in the development of the ARPANET, 1961-1972. **IEEE Annals of the History of Computing**, v. 17, n. 4, p. 76–81, 1995. Disponível em: <https://doi.org/10.1109/85.477437>. Acesso em: 3 nov. 2021.

OBAR, Jonathan A.; OELDORF-HIRSCH, Anne. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. **Information, Communication & Society**, v. 23, n. 1, p. 128-147, 2018.

OECD. **An Introduction to Online Platforms and Their Role in the Digital Transformation**. [S. l.]: OECD, 2018. Disponível em: <https://doi.org/10.1787/53e5f593-en>. Acesso em: 21 mar. 2021.

OECD. **Data-Driven Innovation: Big Data for Growth and Well-Being**. Paris: OECD Publishing, 2015. Disponível em: <https://doi.org/10.1787/9789264229358-en>. Acesso em: 5 set. 2021.

OECD. **Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value**. [S. l.]: OECD Publishing, 2013a. Disponível em: <https://doi.org/10.1787/5k486qtxldmq-en>. Acesso em: 19 set. 2021.

OECD. **Introduction to Data and Analytics (Module 1): Taxonomy, Data Governance Issues and Implications for further Work**. [s. l.], 2013b.

OECD. **Personalised Pricing in the Digital Era**. [S. l.: s. n.], 2018. Disponível em: [https://one.oecd.org/document/DAF/COMP\(2018\)13/en/pdf](https://one.oecd.org/document/DAF/COMP(2018)13/en/pdf). Acesso em: 3 nov. 2021.

OECD. **The Digital Economy**. [s. l.], 2012. Disponível em: <http://www.oecd.org/daf/competition/>. Acesso em: 20 out. 2021.

OFFICE OF FAIR TRADING. **The economics of online personalised pricing Report No. 1488**. [S. l.: s. n.], 2013. Disponível em: <https://docplayer.net/9529539-The-economics-of-online-personalised-pricing.html>. Acesso em: 4 nov. 2021.

OJANPERÄ, Sanna; GRAHAM, Mark; ZOOK, Matthew. The Digital Knowledge Economy Index: Mapping Content Production. **Journal of Development Studies**, v. 55, n. 12, p. 2626–2643, 2019. Disponível em: <https://doi.org/10.1080/00220388.2018.1554208>. Acesso em: 25 ago. 2021.

O QUE É CIÊNCIA DE DADOS? **Oracle**, 2021. Disponível em: <https://www.oracle.com/br/data-science/what-is-data-science/>. Acesso em: 3 nov. 2021.

PFEIFFER, Roberto Augusto Castellanos. Digital Economy, Big Data and Competition Law. **SSRN Electronic Journal**, v. 3, n. 1, p. 53-59, 2019. Disponível em: <https://doi.org/10.2139/ssrn.3440296>. Acesso em: 10 ago. 2021.

PIGOU, Arthur C. **A. C. Pigou. The Economics of Welfare**. Nova Iorque: Macmillan & Co., 1920.

PORTER, Michael E. Competitive advantage: Creating and sustaining superior performance. [S. l.]: 1985.

POSNER, Eric; WEYL, Glen. **Radical Markets: Uprooting Capitalism and Democracy for a Just Society**. Princeton: Princeton University Press, 2018. *E-book*.

PRUFER, Jens; SCHOTTMMLER, Christoph. Competing with Big Data. **SSRN Electronic Journal**, n. 6, 2017. Disponível em: <https://doi.org/10.2139/ssrn.2918726>. Acesso em: 3 nov. 2021.

PWC. **Putting a value on data** . [S. l.], 2019. Disponível em: <https://www.pwc.co.uk/issues/data-analytics/insights/putting-value-on-data.html>. Acesso em: 3 nov. 2021.

RAYPORT, Jeffrey. F.; SVIOKLA, John J. Exploiting the virtual value chain. **Harvard Business Review**, v. 73, n. 6, p. 75-85, 1995. Disponível em: <https://hbr.org/1995/11/exploiting-the-virtual-value-chain>. Acesso em: 3 nov. 2021.

RICHTER, Diogo Kastrup. Desafios e interfaces da defesa da concorrência e da proteção de dados pessoais na economia de plataforma. **Revista de Direito Público da Economia**, v.18, n.72, 2020. Disponível em: <https://dspace.almg.gov.br/handle/11037/39851>. Acesso em: 24 out. 2021.

ROSENBERG, Daniel. Data before the Fact. In: “Raw Data” is an oxymoron. [S. l.]: MIT Press, 2013. p. 15–40.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3. ed. Nova Jersey: Prentice Hall, 2009.

SALUS, Peter H.; BY-CERF, Peter H./Foreword; G., Vinton. **Casting the net: from ARPANET to Internet and beyond**. [S. l.]: Addison-Wesley Pub. Co, 1995. *E-book*.

SANTOS, Babur De los; KOULAYEV, Sergei. Optimizing Click-Through in Online Rankings with Endogenous Search Refinement., **Marketing Science**, v. 36, n. 4, p. 542–564, 2016. Disponível em: <https://doi.org/10.1287/MKSC.2017.1036>. Acesso em: 3 nov. 2021.

SCHWAB, Klaus. **Welcome to the Fourth Industrial Revolution**. Geneva Switzerland: World Economic Forum, 2016.

SETZER, Valdemar W. **Dado, Informação, Conhecimento e Competência**. [S. l.: s. n.], 2001. Disponível em: www.ime.usp.br/~vwsetzer. Acesso em: 24 ago. 2021.

SHAW, Michael J.; SHAPIRO, Carl; VARIAN, Hal R. **Information Rules: A Strategic Guide to the Network Economy**. Massachusetts: Harvard Business Press, 2000.

SHILLER, Ben; WALDFOGEL, Joel. Music for a Song: An Empirical Look at Uniform Pricing and Its Alternatives. **The Journal of Industrial Economics**, v. 59, n. 4, p. 630–660, 2011. Disponível em: <https://doi.org/10.1111/J.1467-6451.2011.00470.X>. Acesso em: 27 out. 2021.

SHILLER, Benjamin Reed. First Degree Price Discrimination Using Big Data. **Working Papers**, [s. l.], 2014. Disponível em: <https://ideas.repec.org/p/brd/wpaper/58.html>. Acesso em: 27 out. 2021.

SHORT, Je; BOHN, Re; BARU, Chaitan. How much information? 2010: Report on enterprise server information. **UCSD Global Information Industry Center**, [s. l.], p. 1–38, 2011. Disponível em: <https://doi.org/10.13140/RG.2.2.22795.21288>. Acesso em: 26 ago. 2021.

STAHL, Florian et al. A classification framework for data marketplaces. **Vietnam Journal of Computer Science**, [s. l.], v. 3, n. 3, p. 137–143, 2016. Disponível em: <https://doi.org/10.1007/s40595-016-0064-2>. Acesso em: 3 nov. 2021.

STIGLER COMMITTEE ON DIGITAL PLATFORMS. **Stigler Committee on Digital Platforms: Final Report - Chicago Booth**. [S. l.: s. n.], 2019. Disponível em: <https://www.chicagobooth.edu/research/stigler/news-and-media/committee-on-digital-platforms-final-report>. Acesso em: 3 nov. 2021.

STROUD, Court. Cambridge Analytica: The Turning Point In The Crisis About Big Data. **Forbes**, 2018. Disponível em: <https://www.forbes.com/sites/courtstroud/2018/04/30/cambridge-analytica-the-turning-point-in-the-crisis-about-big-data/?sh=63c7661348ec#b4fa41b48ecc>. Acesso em: 26 set. 2021.

STUCKE, Maurice E; EZRACHI, Ariel. Looking Up in the Data-Driven Economy. **SSRN Electronic Journal**, n. 333, 2017. Disponível em: <https://doi.org/10.2139/ssrn.2975510>. Acesso em: 27 set. 2021.

SWEDISH BOARD OF COMMERCE. **No Transfer, No Production: A Report on Cross-border Data Transfers, Global Value Chains, and the Production of Goods**. [S. l.: s. n.], 2015. Disponível em: <https://ec.europa.eu/futurium/en/content/no-transfer-no-production-report-cross-border-data-transfers-global-value-chains-and.html>. Acesso em: 3 nov. 2021.

THE ECONOMIST INTELLIGENCE UNIT. **The Business of Data, The Economist Intelligence Unit report**. [S. l.: s. n.], 2015.

TIROLE, Jean. **Competition and the Industrial Challenge for the Digital Age**. [s. l.], 2020. Disponível em: https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/by/tirole/competition_and_the_industrial_challenge_april_3_2020.pdf. Acesso em: 3 nov. 2021.

TIROLE, Jean. **Economics for the common good**. [S. l.]: Princeton University Press, 2017.

TIROLE, Jean. **The Theory of Industrial Organization**. [S. l.]: The MIT Press, 1988. *E-book*.

UBALDI, Barbara. **Open Government Data. Towards Empirical Analysis of Open Government Data Initiatives | Sharing Cities Alliance**. Paris: OECD Publishing, 2013. Disponível em: <https://sharingcitiesalliance.knowledgeowl.com/help/open-government-data-towards-empirical-analysis-of-open-government-data-initiatives>. Acesso em: 3 nov. 2021.

UNCTAD. **Digital Economy Report 2019: Value Creation and Capture**. [S. l.: s. n.], 2019. Disponível em: <https://unctad.org/webflyer/digital-economy-report-2019>. Acesso em: 30 set. 2021.

URSU, Raluca. The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. **SSRN Electronic Journal**, [s. l.], 2016. Disponível em: <https://doi.org/10.2139/SSRN.2729325>. Acesso em: 19 out. 2021.

URSU, Raluca. The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. **Marketing Science**, v. 37, n. 4, p. 530-552, 2016. Disponível em: <https://doi.org/10.2139/SSRN.2729325>. Acesso em: 19 out. 2021.

VAN ALSTYNE, Mashall W.; PARKER, Geoffrey G.; CHOUDARY, Sangeet P. **Pipelines, Platforms, and the New Rules of Strategy**. [S. l.: s. n.], 2016. Disponível em: <https://hbr.org/2016/04/pipelines-platforms-and-the-new-rules-of-strategy>. Acesso em: 3 nov. 2021.

VAN ALSTYNE, Mashall W.; PARKER, Geoffrey G.; CHOUDARY, Sangeet P. **Pipelines, Platforms, and the New Rules of Strategy**. [S. l.: s. n.], 2016. Disponível em: <https://hbr.org/2016/04/pipelines-platforms-and-the-new-rules-of-strategy>. Acesso em: 3 nov. 2021.

VISCONTI, Roberto; LAROCCA, Alberto; MARCONI, Michele. Big Data-Driven Value Chains and Digital Platforms: From Value Co-Creation to Monetization. **SSRN Electronic Journal**, 2017. Disponível em: <https://doi.org/10.2139/ssrn.2903799>. Acesso em: 6 set. 2021.

WEBER, Steven. Data, development, and growth. **Business and Politics**, v. 19, n. 3, p. 397-423, 2017. Disponível em: <https://doi.org/10.1017/BAP.2017.3>. Acesso em: 3 nov. 2021.

WEINBERGER, David. Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room. [S. l.]: Basic Books, 2011.

WORLD DEVELOPMENT REPORT, 2016: DIGITAL DIVIDENDS. **Choice Reviews Online**, [s. l.], v. 53, n. 11, p. 53-4889-53-4889, 2016. Disponível em: <https://doi.org/10.5860/choice.196952s>. Acesso em: 3 nov. 2021.

YAMANA, Rieti; NIREI, Makoto; YAMANA, Kazufumi. Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy. **RIETI Discussion Paper Series**, [s. l.], p. 19-022, 2019. Disponível em: <https://www.rieti.go.jp/en/publications/summary/19030027.html>. Acesso em: 20 mar. 2021.

YULINSKY, Corey. Decisions, decisions ... will 'Big Data' have 'Big' impact? | Financial Times. **Financial Times**, [s. l.], 24 jan. 2012. Disponível em: <https://www.ft.com/content/9ee048b6-4612-11e1-9592-00144feabdc0>. Acesso em: 3 nov. 2021.

ZHU, Feng et al. Network Interconnectivity and Entry into Platform Markets. **Working Paper 19-062**, [s. l.], 2019.