Maciel: Data scraping e responsabilidade do controlador

Os vazamentos de dados pessoais têm se tornado cada vez mais frequentes. Uns atribuem ao "efeito LGPD", uma vez que os criminosos se aproveitam das graves consequências reputacionais e pecuniárias para extorquirem as organizações e, quando essas não cedem, divulgam os dados por ela controlados. Outros, por suas vezes, arguem que o aumento das ocorrências se dá pelo elevado valor dessas

Em 2021, tivemos no Brasil o "vazamento do fim do mundo" ou "megavazamento", como ficou chamada a divulgação de base de dados de mais de 233 milhões de brasileiros, incluídos falecidos. Recentemente, dois novos episódios globais vieram à tona: Facebook [1] e Clubhouse [2]. Pelo primeiro, 533 milhões de pessoas foram afetadas, sendo oito milhões brasileiros, e, pelo segundo, 1,3 milhão no total, não se sabendo ao certo quantos brasileiros vitimados. Em nenhum desses casos é possível afirmar a ocorrência de vazamento a partir de ativos das respectivas empresas. Estão sob escrutínio das autoridades.

No dito "megavazamento", pelos tipos de dados divulgados, a origem foi creditada à Serasa, que sempre negou qualquer intercorrência em seus sistemas. Cogita-se, inclusive, ter havido uma compilação de diversas bases de dados para garantir um maior volume e, consequentemente, agregar valor ao preço de comercialização, ou, como dito por um dos *hackers* presos pela PF, teria sido originado de outra empresa [3]. As redes sociais, por seu turno, também insistem não ter ocorrido vazamento e que os dados eram públicos e, por assim serem, provavelmente foram extraídos por APIs ou por métodos de *data scraping*. *Data Scraping* é um termo geral que descreve uma miríade de possibilidades de metodologias de extração de dados, com ou sem permissão do titular dos dados. Não é ilícito por si só, tanto que APIs são utilizadas com frequência pelos usuários para, por exemplo, cadastrarem em serviços de terceiros.

Importante relembrar que a informação deixada pública pelo usuário em uma rede social não representa uma autorização para utilizá-la em outra finalidade. Ao contrário do que uns podem pensar, a previsão do §4º do artigo 7º da LGPD de que é dispensada a exigência do consentimento para os dados tornados manifestamente públicos pelo titular não é um passaporte para ampla e irrestrita utilização. Tanto que o referido dispositivo faz a prudente ressalva: *"Resguardados os direitos do titular e os princípios previstos nesta lei"*.

O tratamento de dados pessoais deve sempre pautar-se pelo princípio da finalidade, ou seja, não se pode desvirtuar o tratamento para finalidades distintas daquelas previamente informadas ao titular dos dados. Se o usuário deixa suas informações públicas na rede social, seja para que o encontrem mais facilmente

ou mesmo para divulgar suas opiniões e imagens, a publicidade não pode servir de argumento para usos diversos, sobretudo fora daquele ambiente.

A ministra Rosa Weber do Supremo Tribunal Federal em julgamento [4] para suspender a Medida Provisória nº 954/2020, que visava à transferência de dados pessoais de empresas de telecomunicação à Fundação IBGE, fundamentou na época — quando sequer a LGPD estava vigente — que o caso tratava-se de proteção constitucional à privacidade, citando o direito à autodeterminação informativa expressamente, como também seus correlacionados princípios da finalidade, necessidade e adequação. *Verbi gratia:*

"A Constituição da República confere especial proteção à intimidade, à vida privada, à honra e à imagem das pessoas ao qualificá-las como invioláveis, enquanto direitos fundamentais da personalidade, assegurando indenização pelo dano material ou moral decorrente de sua violação (artigo 5º, X). O assim chamado direito à privacidade (right to privacy) e os seus consectários direitos à intimidade, à honra e à imagem emanam do reconhecimento de que a personalidade individual merece ser protegida em todas as suas manifestações.
(…)
***Decorrências dos direitos da personalidade, o respeito à privacidade e à autodeterminação informativa** foram positivados, no artigo 2º, I e II, da Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais), como fundamentos específicos da disciplina da proteção de dados pessoais.*
(…)
*Nessa linha, ao não definir apropriadamente como e para que serão utilizados os dados coletados, a MP n. 954/2020 **não oferece condições para avaliação da sua adequação e necessidade, assim entendidas como a compatibilidade do tratamento com as finalidades informadas e sua limitação ao mínimo necessário para alcançar suas finalidades**. Desatende, assim, a garantia do devido processo legal (artigo 5º, LIV, da Lei Maior), em sua dimensão substantiva"* (grifos do autor).

Nessa mesma esteira, pouco antes, o Tribunal da Cidadania, em julgamento de caso envolvendo o compartilhamento de dados de consumo a órgãos de proteção ao crédito, chancelou importante passagem, sob uso de dados expostos publicamente pelo próprio titular. Constou da ementa [5]:

"9. O fato, por si só, de se tratarem de dados usualmente fornecidos pelos próprios consumidores quando da realização de qualquer compra no comércio, não afasta a responsabilidade do gestor do banco de dados, na medida em que, quando o consumidor o faz não está, implícita e automaticamente, autorizando o comerciante a divulgá-los no mercado;
10. Do mesmo modo, o fato de alguém publicar em rede social uma informação de caráter pessoal não implica o consentimento, aos usuários que acessam o conteúdo, de utilização de seus dados para qualquer outra finalidade, ainda mais com fins lucrativos".

Vale lembrar também ter sido justamente o uso indevido por terceiros de dados coletados no Facebook que provocou uma das maiores polêmicas relacionadas à privacidade na era da informação: o "caso Cambridge Analytica". O uso indevido dos dados pessoais dos usuários da rede social por essa empresa para finalidades diversas, sem informação ou consentimento eficaz dos usuários acarretou ao Facebook, por ter sido conivente, a maior multa já aplicada por violação de dados pessoais no mundo: US$ 5 bilhões aplicada pela Federal Trade Commission (FTC) [6], órgão fiscalizatório dos Estados Unidos da América. Sem prejuízo das demais multas aplicadas por autoridades de proteção de dados de outros países, como a ICO no Reino Unido, que fixou em 500 mil libras esterlinas, equivalente a

aproximadamente R$ 3,5 milhões, maior valor permitido para sanção naquele país.

Nesse viés, temos que aqueles que utilizam os dados de forma desvirtuada, sem base legítima — seja de forma manual, pontualmente ou automática, a partir de robôs de extração (*data scraping*) —, cometem ilícito e respondem perante os titulares de dados e autoridades fiscalizatórias. Entretanto, por ora, não aprofundaremos o estudo nesse sentido.

Seguiremos por uma outra trilha, com enfoque não em quem extraiu e desvirtuou a finalidade, mas, sim, na seguinte *quaestio juris*: Há responsabilidade do controlador que permitiu a extração dos dados de sua plataforma? São o Facebook e Clubhouse responsáveis por não bloquear esse mecanismo?

O artigo 42 da LGPD estabelece que tanto o controlador como o operador respondem pelos danos patrimoniais, morais, individuais ou coletivos quando realizarem tratamento de dados de forma ilícita.

Por sua vez, o tratamento é claramente ilegal quando não puder fornecer a segurança que o titular dele pode esperar (artigo 44), considerados: o modo pelo qual é realizado; o resultado e os riscos que razoavelmente dele se esperam; e as técnicas de tratamento de dados pessoais disponíveis à época em que foi realizado. O parágrafo único acrescenta que os agentes de tratamento também responderão quando não adotarem as medidas técnicas e administrativas aptas a protegerem os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito (artigo 46).

Segurança é um pilar fundamental no tratamento de dados pessoais para assegurar e fortalecer a confiança do titular perante o agente de tratamento. Não se pode ignorar que, ao permitir a extração de dados de suas plataformas por ferramentas automatizadas e em larga escala, sem permissão, o agente de tratamento falha com seu dever de proteger os dados contra tratamento inadequado ou ilícito. E, frisamos, é ilícito usar dados para finalidades diversas daquelas para os quais foram inicialmente obtidos.

Não importa que a ilicitude desse tratamento seja cometida por terceiro. Uma vez não tendo o agente tomado as medidas cabíveis e disponíveis à época para evitar esse tratamento indevido responde pela reparação dos danos e demais sanções previstas na lei.

Não por acaso que a lei expressou claramente que as medidas de segurança devem ser pensadas desde a sua concepção (§3º, artigo 46) — *privacy by design*. É inútil expor em suas políticas, como o fez a rede social Clubhouse, que não permite o acesso indevido, se o agente não adotar medidas efetivas em seus sistemas que bloqueassem a extração.

De acordo com o pesquisador de segurança da informação da CyberNews Mantas Sasnauskas, há falhas no aplicativo do Clubhouse que permitem a qualquer um com apenas um token ou via uma API extrair todas as informações públicas dos usuários, e referido token sequer expira. O pesquisador complementa que *"não ter uma medida de anti-scraping implementada é uma questão de privacidade"* [7]
. Complementamos: violação de segurança e quebra da confiança dos usuários, que deixaram sua informação pública para as finalidades de uso da rede social e não para servirem de formação de perfis comportamentais ou vendas criminosas pelos becos da internet.

Posteriormente à enxurrada de ações judiciais e o escrutínio público após o "caso Cambridge Analytica", o Facebook encrudesceu suas políticas contra scraping desautorizado, dedicando um time de mais de

cem pessoas (*External Data Misuse* — EDM), incluindo cientistas de dados e engenheiros focados em "detectar, bloquear e deter" *scraping* [8]. Adotou, por exemplo, limitações na taxa e volume de dados para a extração de dados, além de ações judiciais contra scrapers recorrentes.

O esforço é válido, porém ainda muito tímido para impedir a utilização de *scraping* desautorizado. Mas sinaliza o óbvio e o previsto na lei, que é pensar em proteger a privacidade em todas as suas vertentes de forma *"proativa e não reativa; preventiva, não corretiva"*, tal como estatuído nos sete princípios do *privacy by Design* da lavra de Ann Cavoukian [9].

Demonstrada a não adoção de medidas efetivas pelo agente de tratamento — considerados seu porte econômico, volume de dados e tecnologia atual —, não há dúvida de que devem responder pelos danos causados aos titulares de dados.

Como já se tem visto por aí, nem há de se falar em excludente de responsabilidade por culpa exclusiva do titular dos dados calçado no argumento de que o usuário deixou seus dados públicos correndo o risco de uso indevido por terceiros, porquanto não se aplicam quaisquer um dos requisitos previstos no artigo 43, mormente por não ser a culpa exclusiva do titular.

Ainda que os dados sejam públicos, permitir a raspagem de dados desautorizada é uma falha de segurança. Quando muito deve se permitir tais métodos mediante supervisão e para finalidades legítimas, também impondo limites para que, mesmo sob tais condições, apenas sejam tratados os dados estritamente necessários.

[1] https://www.reuters.com/article/us-facebook-data-leak-idUSKBN2BU2ZY. Acessado em 15/4/2021.

[2] https://olhardigital.com.br/2021/04/12/seguranca/clubhouse-vazamento-expoe-dados-de-13-milhao-de-usuarios

[3] https://thehack.com.br/seria-o-maior-vazamento-de-dados-do-brasil-uma-fraude-especialistas-comentam/. Acessado em 15/4/2021

[4] STF, medida cautelar na ação direta de inconstitucionalidade 6.387 Distrito Federal.

[5] STJ, Recurso especial nº 1.758.799 / MG

[6] https://www.wsj.com/articles/facebooks-5-billion-privacy-settlement-wins-court-approval-11587752759

[7] https://cybernews.com/security/clubhouse-data-leak-1-3-million-user-records-leaked-for-free-online/. Acessado em 15/4/2021.

[8] https://about.fb.com/news/2021/04/how-we-combat-scraping

[9] https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf. Acessado em 19/4/2021.

Date Created
05/05/2021