

Efeitos da GDPR no Trabalho de Cientistas dos Dados

A recente vigoração da *General Data Protection Regulation* (GDPR) impõe restrições sobre a coleta e processamento de dados pessoais, atribuindo certas responsabilidades a cientistas dos dados quanto a seus trabalhos. Apesar de se tratar de uma regulamentação aplicada à União Europeia e o Reino Unido, seus efeitos se sentem por toda a rede, visto que as principais empresas afetadas se trata de multinacionais. Com mais intensidade do que antes, a tarefa de processamento deve ser realizada com grande cautela direcionada à proteção da privacidade dos indivíduos sujeitos à coleta de dados e à justiça em decisões que os afetem.

Dados coletados devem ser somente os pertinentes à tarefa desejada, a fim de evitar decisões enviesadas. O uso de dados como idade, raça e etnia para decisões automatizadas é estritamente proibido. Não somente deve-se excluir esses dados dos conjuntos usados, mas é preciso garantir que os dados presentes nesses conjuntos não tragam um viés relacionado a esses tipos de dados, seja por colinearidade ou outra representatividade. Se necessário, deve-se observar toda a linhagem dos dados, já que derivações sucessivas dificultam determinar erros ao longo do processo.

É importante atentar também para a integridade dos dados utilizados, para que sejam representantes válidos da realidade a ser modelada e não prejudiquem a grupos específicos em suas decisões. Isso reforça a importância da análise exploratória dos dados e maior participação na sua aquisição. Além dessas condições, é mandatório expor aos consumidores que seus dados estão sendo coletados, bem como esclarecer que serão utilizados para tomadas de decisão automatizadas.

Passada a fase de aquisição e análise de dados, vem os cuidados quanto às formas de processamento dos dados. A GDPR assegura a consumidores o que se pode chamar de “direito a explicação”, de forma que, em situações de decisão automatizada sobre certo indivíduo, como a concessão ou não de crédito, promoção ou não em seu emprego, entre outras que o afetem legalmente, explicações ou descrições dos processos utilizados para levar a tal decisão devem ser fornecidas, caso solicitadas. Há controvérsias sobre as reais implicações desse direito por sua descrição ambígua, e seus efeitos ainda hão de ser vistos em casos vindouros. Caso não bastem descrições sobre os algoritmos utilizados para tomadas de decisões automáticas, e sejam necessárias explicações detalhadas para consumidores dos motivos por quais obtiveram suas decisões, como, por exemplo, detalhar para um cliente por que teve seu crédito negado com base na valoração do conjunto de variáveis, entre *credit scores* e histórico de compras, algoritmos dos quais não se podem extrair explicações claras podem tornar-se ilegais, como o *Deep Learning*.

O dito “direito a explicação” preocupa empresas também por poder forçá-las a revelar seus algoritmos de automação de decisão, que em segredo podem representar vantagem competitiva, mas o contrário caso expostos.

Cabe esclarecer que em casos que dados de consumidores forem usados com suficiente anonimidade, para tomada de decisões não sobre um consumidor específico mas de forma ampla, não é direito do consumidor solicitar explicações sobre o funcionamento do algoritmo de decisão. Como exemplo, pode-se citar o algoritmo de recomendação de vídeos do YouTube, que utiliza dados de visualização de vídeos de seus usuários.

Porém, existem otimistas do outro lado da ambiguidade. Alguns argumentam que leis semelhantes já existem em muitas regiões da Europa, sem carregar as temidas consequências que se supõem para a GDPR. A exemplo, há no Reino Unido regulações que permitem a consumidores solicitar explicações caso recebam decisões adversas de concessão de crédito. Há argumentos também de que para situações em que explicações possam ser solicitadas, algoritmos comumente usados são justamente os que permitem que se extraia deles informações de como dados de entrada influenciam suas decisões, por exemplo, *Decision Tree*.