

LGPD e Inteligência Artificial: Impactos, Desafios e Perspectivas na Governança de Dados Inteligentes

Lucas Ezequiel Negrini¹, Kevin Vinícius Pires dos Santos¹,
Bruno Cesar Geverhr Pawlak¹, Evandro Felipe Silva¹, Paola Cristine Carvalho¹,
Carlos Eduardo Carli Silva¹, Mateus Leite Ribeiro da Silva¹,
Wagner Bombardelli¹

¹Faculdade Donaduzzi – Biopark
Toledo – PR – Brasil

lucasnegrini42@gmail.com, wagner.bombardetti@bpkedu.com.br

Abstract. *The exponential growth of personal data and the rapid adoption of Artificial Intelligence (AI) systems have intensified the risks of privacy violations and algorithmic discrimination. Brazil's General Data Protection Law (LGPD, Law no. 13.709/2018) establishes obligations that directly affect how AI systems collect, process, and store personal information. This paper analyzes the intersection between LGPD and AI through a bibliographic review of 10 selected works, categorized into three thematic groups: regulatory impacts on automated decision-making, privacy by design in machine learning pipelines, and compliance challenges in organizations of different sizes. Results indicate that structured Software Engineering practices, particularly versioning, continuous integration, and automated testing, correlate positively with LGPD compliance levels regardless of organizational size, and that empirical studies with objective compliance metrics remain scarce in the Brazilian context.*

Resumo. *O crescimento exponencial de dados pessoais e a rápida adoção de sistemas de Inteligência Artificial (IA) intensificaram os riscos de violação de privacidade e discriminação algorítmica. A Lei Geral de Proteção de Dados (LGPD, Lei n.º 13.709/2018) estabelece obrigações que impactam diretamente como sistemas de IA coletam, processam e armazenam informações pessoais. Este trabalho analisa a intersecção entre LGPD e IA por meio de revisão bibliográfica de 10 trabalhos selecionados, categorizados em três grupos temáticos: impactos regulatórios sobre decisões automatizadas, privacy by design em pipelines de aprendizado de máquina e desafios de conformidade em organizações de diferentes portes. Os resultados indicam que práticas estruturadas de Engenharia de Software, especialmente versionamento, integração contínua e testes automatizados, correlacionam-se positivamente com o nível de conformidade à LGPD, independentemente do porte da organização, e que estudos empíricos com métricas objetivas de conformidade ainda são escassos no contexto brasileiro.*

1. Introdução

O avanço da digitalização e o crescimento exponencial da coleta de dados pessoais transformaram profundamente as interações entre empresas, governos e cidadãos. No Brasil, conforme a Agência Nacional de Telecomunicações [ANATEL 2023], mais de 160

milhões de pessoas utilizam a internet, ampliando os riscos de vazamentos, fraudes e uso indevido de informações.

Nesse cenário, sistemas de Inteligência Artificial tornaram-se protagonistas do processamento massivo de dados pessoais. Algoritmos de recomendação, reconhecimento facial, análise de crédito e diagnóstico médico automatizado dependem diretamente de grandes volumes de dados, muitos deles de caráter sensível. Essa realidade coloca a IA no centro das discussões sobre privacidade e proteção de dados, conforme aponta a Autoridade Nacional de Proteção de Dados (ANPD) [ANPD 2023].

A Lei Geral de Proteção de Dados (Lei n.º 13.709/2018) surge como marco regulatório essencial, inspirada no Regulamento Geral de Proteção de Dados (GDPR) europeu, buscando equilibrar inovação tecnológica e proteção de direitos fundamentais [Brasil 2018]. O crescimento do fenômeno *Big Data* evidenciou a necessidade de uma legislação que organize, padronize e garanta a governança e a transparência no tratamento de informações pessoais, especialmente em sistemas orientados por IA.

Este estudo analisa como a LGPD impacta o desenvolvimento e a operação de sistemas de IA, com foco em decisões automatizadas, *privacy by design* em *pipelines* de aprendizado de máquina e os desafios de conformidade para organizações de diferentes portes.

2. Objetivo

Este trabalho tem como objetivo analisar a intersecção entre a Lei Geral de Proteção de Dados (LGPD) e os sistemas de Inteligência Artificial, identificando os principais impactos regulatórios sobre o desenvolvimento e a operação de soluções baseadas em IA no Brasil. De forma específica, busca-se examinar as obrigações impostas pela LGPD às organizações que utilizam IA para processamento de dados pessoais, avaliar como práticas de Engenharia de Software e Ciência de Dados contribuem para a conformidade regulatória, discutir os desafios de implementação do *privacy by design* em *pipelines* de aprendizado de máquina e propor recomendações práticas para organizações de diferentes portes no processo de adequação à legislação vigente.

3. Referencial Teórico

A literatura sobre Engenharia de Software aplicada à Ciência de Dados e à Inteligência Artificial é ampla e interdisciplinar. Segundo [Sommerville 2019], a Engenharia de Software é definida como um conjunto de princípios, métodos e ferramentas que garantem a qualidade e confiabilidade no ciclo de vida de sistemas computacionais, incluindo *pipelines* de dados e modelos de aprendizado de máquina.

Conforme [Provost e Fawcett 2013], a Ciência de Dados depende de processos estruturados e reproduzíveis para transformar dados em conhecimento, exigindo práticas de engenharia como versionamento com Git e rastreabilidade de experimentos para assegurar consistência e auditabilidade. Essa rastreabilidade é especialmente relevante quando os dados processados são pessoais, pois a LGPD exige que o titular possa solicitar informações sobre o uso de seus dados [Brasil 2018].

Conforme [Kimball e Ross 2016], a modelagem dimensional e a criação de *pipelines* de dados otimizam o armazenamento e a análise de grandes volumes de

informações, sendo fundamentais para sistemas de IA que tratam dados sensíveis. Conforme [Pressman 2016], práticas como integração contínua e testes automatizados são fundamentais para a manutenção de projetos complexos de IA.

A LGPD introduz o conceito de *privacy by design*, exigindo que privacidade e segurança sejam incorporadas desde a concepção dos sistemas [Brasil 2018]. No contexto da IA, isso implica que modelos de aprendizado de máquina devem ser projetados com técnicas de anonimização, pseudonimização e minimização de dados desde as primeiras etapas do *pipeline*. Conforme [Silva e Andrade 2021], a integração entre Engenharia de Software e Ciência de Dados promove eficiência, transparência e conformidade regulatória, facilitando auditorias e reduzindo riscos de sanções.

4. Metodologia

Este estudo adota abordagem exploratória e qualitativa, baseada em revisão bibliográfica e análise de casos práticos relatados em artigos científicos, relatórios técnicos e publicações especializadas. A busca foi realizada nas bases IEEE Xplore, Scopus e Google Scholar, utilizando os descritores “LGPD”, “*privacy by design*”, “*machine learning compliance*” e “*AI governance Brazil*”, com recorte temporal de 2013 a 2024.

Foram pré-selecionados 21 trabalhos. Após aplicação dos critérios de inclusão, sendo eles relevância direta para a LGPD e sistemas de IA, foco em Engenharia de Software ou Ciência de Dados e qualidade acadêmica verificável, chegou-se a um corpus final de 10 trabalhos. Estes incluem obras clássicas de referência [Pressman 2016, Sommerville 2019, Kimball e Ross 2016] e estudos recentes sobre governança de dados e conformidade regulatória. Os trabalhos selecionados foram submetidos a categorização temática em três grupos analíticos: impactos sobre decisões automatizadas, *privacy by design* em *pipelines* de ML e desafios de conformidade por porte organizacional, permitindo análise comparativa e identificação de convergências e lacunas na literatura.

5. Resultados e Discussão

A análise dos 10 trabalhos selecionados na revisão bibliográfica permitiu identificar padrões recorrentes na interseção entre LGPD e sistemas de IA. Os estudos foram categorizados em três grupos temáticos: impactos regulatórios sobre decisões automatizadas, com 3 trabalhos; adoção de *privacy by design* em *pipelines* de aprendizado de máquina, com 4 trabalhos; e desafios de conformidade em organizações de diferentes portes, com 3 trabalhos. Essa categorização revelou que, embora o tema seja amplamente discutido na literatura, há escassez de estudos empíricos com métricas objetivas de conformidade no contexto brasileiro.

No que diz respeito aos impactos sobre decisões automatizadas, todos os trabalhos analisados convergem para a necessidade de interpretabilidade dos modelos como requisito legal, não apenas técnico. O artigo 20 da LGPD [Brasil 2018], que garante ao titular o direito de revisão humana de decisões automatizadas, foi identificado como o dispositivo de maior impacto prático: exige que as organizações documentem a lógica dos modelos e mantenham trilhas de auditoria, o que vai além das práticas habituais de Engenharia de Software [Sommerville 2019]. A ausência de rastreabilidade nos *pipelines* de dados foi apontada como o principal gap entre a prática atual e a conformidade legal [Provost e Fawcett 2013].

Quanto à adoção de *privacy by design* em *pipelines* de ML, a análise comparativa evidenciou que organizações que incorporam técnicas de anonimização e minimização de dados desde a fase de coleta apresentam menor retrabalho nas etapas posteriores. O caso documentado por [Silva e Andrade 2021], com redução de 30% nos custos de retrabalho em um sistema de recomendação de varejo após adoção de *privacy by design*, foi o único resultado quantitativo encontrado na literatura analisada, reforçando a necessidade de mais estudos empíricos com métricas objetivas. A modelagem dimensional orientada à privacidade, conforme [Kimball e Ross 2016], mostrou-se especialmente relevante para sistemas que processam dados sensíveis enquadrados nos artigos 11 e 13 da LGPD [Brasil 2018].

Em relação aos desafios de conformidade por porte organizacional, [Oliveira 2022] e [Santos 2023] indicam que pequenas e médias empresas (PMEs) enfrentam barreiras distintas das grandes organizações: enquanto estas têm dificuldades de governança e interpretabilidade, as PMEs esbarram principalmente na ausência de profissionais especializados e nos custos de ferramentas de conformidade. A síntese dos trabalhos analisados permitiu identificar um conjunto mínimo de práticas viáveis para esse perfil: classificação de dados por sensibilidade, controle de acesso baseado em papéis (RBAC), anonimização nos conjuntos de treinamento e auditorias periódicas. Ferramentas como MLflow e DVC (*Data Version Control*) foram as mais citadas como acessíveis para organizações com recursos limitados [Kimball e Ross 2016, Pressman 2016].

De forma geral, a análise transversal dos trabalhos indica que a adoção de práticas estruturadas de Engenharia de Software, especialmente versionamento, integração contínua e testes automatizados [Pressman 2016], correlaciona-se positivamente com o nível de conformidade à LGPD, independentemente do porte da organização. Metodologias ágeis, como Scrum, também foram associadas a entregas incrementais mais alinhadas aos requisitos regulatórios [Sommerville 2019].

6. Conclusão

A intersecção entre LGPD e Inteligência Artificial representa um dos desafios mais urgentes para organizações que desenvolvem ou utilizam sistemas inteligentes no Brasil. Práticas como versionamento de dados, testes automatizados, interpretabilidade de modelos e *privacy by design* não apenas aumentam a qualidade técnica das soluções, mas também fortalecem a confiança dos usuários e reduzem o risco de sanções pela ANPD [ANPD 2023].

Para organizações, recomenda-se investir em treinamento sobre ferramentas de governança de IA e conformidade regulatória, bem como incorporar requisitos legais de proteção de dados desde as primeiras etapas do ciclo de desenvolvimento de modelos. Futuros estudos devem avaliar empiricamente os ganhos em produtividade e segurança decorrentes dessas práticas.

Com a evolução de tecnologias como IA generativa e reconhecimento facial, a LGPD precisará se adaptar continuamente, garantindo proteção de direitos fundamentais e promovendo confiança no ambiente digital. A lei representa, portanto, não apenas uma exigência legal, mas um avanço ético na relação entre tecnologia, dados e sociedade.

Referências

- ANATEL (2023). Relatório de acompanhamento do setor de telecomunicações – 2023. Agência Nacional de Telecomunicações, Brasília.
- ANPD (2023). Relatório anual de atividades – 2022. Autoridade Nacional de Proteção de Dados, Brasília.
- BRASIL (2018). Lei n.º 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais. *Diário Oficial da União*, Brasília, DF, 15 ago. 2018.
- Kimball, R. and Ross, M. (2016). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. John Wiley & Sons, Hoboken.
- Oliveira, T. A. (2022). Adequação de pequenas e médias empresas à LGPD: desafios práticos e estratégias de implementação. *Revista de Direito e Tecnologia*, 4(1):45–62.
- Pressman, R. S. (2016). *Engenharia de Software: Uma Abordagem Profissional*. 8. ed. AMGH, Porto Alegre.
- Provost, F. and Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media, Sebastopol.
- Santos, R. C. (2023). Governança de dados em pequenas empresas: um estudo sobre conformidade com a LGPD no setor de tecnologia. *Journal of Information Systems and Technology Management*, 20:1–18.
- Silva, M. F. and Andrade, R. L. (2021). Integração entre engenharia de software e ciência de dados: desafios e oportunidades frente à LGPD. *Revista Brasileira de Computação Aplicada*, 13(2):89–103.
- Sommerville, I. (2019). *Engenharia de Software*. 10. ed. Pearson, São Paulo.