
Risk-Gov-AI¹

Elaine Fabíola Soares²
Track 1: Latin America-Governance

With
Apart Research

Abstract

This work presents RiskGovAI, a platform developed during the Global South AI Safety Hackathon to support Artificial Intelligence governance and the consultation of data protection legislation in Latin America. The solution aims to automate regulatory research, reduce interpretation errors, and facilitate access to normative information. The architecture utilizes the Gemma 3 4B language model integrated with the Retrieval-Augmented Generation (RAG) technique, allowing for contextualized queries based on official documents. For information storage and retrieval, ChromaDB was employed—a vector database responsible for indexing and retrieving relevant excerpts from the analyzed regulations. The knowledge base brings together Brazil's General Data Protection Law (LGPD), Chile's Law 21.719, Colombia's Law 1581, and Mexico's Federal Law on the Protection of Personal Data Held by Private Parties (LFPDPPP). Beyond handling queries, the platform allows for the comparison of requirements across different jurisdictions, identifying regulatory similarities and differences. To increase response reliability and reduce hallucinations, supervision mechanisms known as Guardrails were implemented, ensuring greater alignment with the source materials used. The results demonstrate that the combination of LLMs, RAG, ChromaDB, and Guardrails can expand access to regulatory knowledge and support responsible AI governance practices, contributing to safety and reliability goals.

¹ Research conducted at the [The Global South AI Safety Hackathon](#), June 2026

² Federal University of São Carlos - Computer Science - elaine.soares@estudante.ufscar.br

1. Introduction

The advancement of Artificial Intelligence (AI) brings economic benefits, but imposes challenges regarding privacy, security, and accountability (OECD, 2019; UNESCO, 2021). In Latin America, this scenario is complex due to regulatory heterogeneity across countries, which hinders compliance for developers and organizations (UNESCO, 2021).

Although the region has strengthened its regulatory frameworks, such as the LGPD in Brazil (2018), Law 21.719 in Chile (2024), Law 1581 in Colombia (2012), and the LFPDPPP in Mexico (2010), the manual and comparative analysis of these legislations is time-consuming. It requires high specialization, and increases the risk of interpretative errors.

To mitigate this bottleneck, this work presents RiskGovAI, a platform designed to automate regulatory research, mitigate interpretation errors, and democratize access to data protection regulatory information in Latin America.

2. Architecture and Technical Foundations

To ensure factual accuracy, RiskGovAI utilizes an approach based on prompt engineering and vector storage, structured around four pillars:

- **Gemma 3 4B:** An open language model that uses local attention layers to reduce memory usage (KV-cache) in long contexts (up to 128K tokens). Trained via distillation, it delivers high performance in legal reasoning and instruction-following in a computationally lightweight manner.
- **RAG (Retrieval-Augmented Generation):** A technique that connects the model to an external memory (Lewis et al., 2020). Instead of relying on the LLM's static knowledge, the system retrieves relevant excerpts from the laws before generating the response, ensuring factual grounding.
- **ChromaDB:** An open-source vector database (Apache 2.0) responsible for storing the segmented text (chunks) of the LGPD, Law 21.719, Law 1581, and LFPDPPP converted into embeddings. It performs hybrid searches based on semantic similarity and applies metadata filters for comparative analyses.
- **Guardrails:** Supervision mechanisms that validate user input and model output. They monitor the alignment of the generated text with ChromaDB sources, mitigating hallucinations and ensuring the reliability required in compliance audits.

3. Related Work

The integration of LLMs and RAG in the legal domain has expanded significantly since Lewis et al. (2020), with benchmarks like **LegalBench** (Guha et al., 2023) proving that generic models can handle complex compliance when paired with specialized knowledge bases. However, current state-of-the-art tools focus almost exclusively on US/UK common law or European frameworks (GDPR/EU AI Act), leaving Latin American legislations underrepresented. Furthermore, existing legal QA systems are usually siloed, lacking cross-border comparative capabilities.

RiskGovAI addresses these gaps through three main contributions:

- **Cross-Border Harmonization:** It uses ChromaDB metadata filtering to enable multi-jurisdiction comparative queries, mapping regulatory differences across Brazil, Chile, Colombia, and Mexico.
- **Resource-Efficient Open Architecture:** By deploying **Gemma 3 4B**, the platform leverages optimized local attention to process long legal contexts (128K tokens) with low memory overhead, making it viable for the Global South's hardware constraints.
- **Double-Layered Reliability:** Unlike raw LLM pipelines, it couples RAG with strict supervisory Guardrails to eliminate hallucinations and guarantee the deterministic alignment required for legal compliance.

4. Methods

4.1 System Architecture

RiskGovAI was developed to support AI governance and regulatory compliance analysis across Latin America. The platform combines the Gemma 3 4B large language model, Retrieval-Augmented Generation (RAG), ChromaDB, and safety mechanisms (Guardrails) to provide answers grounded in official legal and regulatory documents.

The workflow consists of four main stages:

1. Collection and preparation of regulatory documents.
2. Translation and preprocessing of legal texts.
3. Embedding generation and indexing in a vector database.
4. Context retrieval and response generation.

4.2 Data Collection and Processing

The knowledge base was built using official regulatory documents related to data protection and AI governance from four Latin American jurisdictions:

- Brazil: General Data Protection Law (LGPD – Law No. 13,709/2018);
- Chile: Law No. 21.719;
- Colombia: Law No. 1581/2012;

- Mexico: Federal Law on the Protection of Personal Data Held by Private Parties (LFPDPPP).

The legal documents were initially collected in PDF format from official government sources. To facilitate processing within the RAG pipeline, the files were converted into Markdown (MD) format while preserving their textual structure and content.

After conversion, all documents were translated into English using Gemini 2.5 Pro. This decision was motivated by the stronger performance of modern language models and embedding models when processing English-language content. Standardizing documents in a single language also improved consistency across different jurisdictions, enabling more accurate semantic retrieval and comparison of legal requirements.

The translated documents were then segmented into smaller text chunks before being indexed in the vector database.

4.3 Embeddings and Information Retrieval

RiskGovAI uses vector embeddings stored in ChromaDB to perform semantic search across legal documents. Unlike traditional keyword-based systems, embeddings capture the contextual meaning of text, allowing the platform to identify conceptually related information even when different terminology is used.

This capability is particularly important in the legal domain, where similar regulatory concepts may be expressed differently across jurisdictions. For example, terms such as “sensitive personal data,” “special categories of personal data,” or equivalent legal expressions can be associated through semantic similarity despite variations in wording. According to Tamkin et al. (2021), contextual embeddings capture semantic relationships based on surrounding text, reducing ambiguities in language interpretation.

When a user submits a query, the system converts the request into an embedding and performs a similarity search in ChromaDB. The most relevant document chunks are retrieved and provided as context to the Gemma 3 4B model. Following the Retrieval-Augmented Generation approach proposed by Lewis et al. (2020), this process grounds responses in authoritative sources and improves factual accuracy.

4.4 Safety and Reliability

To increase reliability and mitigate hallucinations, RiskGovAI incorporates Guardrails that validate user inputs and constrain model outputs to information supported by retrieved documents. These safeguards ensure that responses remain aligned with the regulatory sources stored in the knowledge base.

Additionally, the system provides references to the retrieved legal passages, improving transparency and enabling users to verify the origin of the information presented.

4.5 Deployment and Design Decisions

The prototype was deployed on a low-cost Virtual Private Server (VPS) to demonstrate the feasibility of implementing AI governance solutions without requiring specialized infrastructure. The deployment environment consisted of:

- 2 vCPU cores;
- 8 GB RAM;
- 100 GB NVMe storage;
- 8 TB monthly bandwidth.

These hardware specifications were intentionally selected to validate that the proposed architecture could operate on affordable cloud infrastructure accessible to small organizations, researchers, and public institutions in the Global South.

The Gemma 3 4B model was chosen because it provides a balance between performance and computational efficiency, allowing deployment within the available hardware constraints. Larger models were considered but were not adopted due to their higher memory and processing requirements.

Initial experiments using only the language model resulted in less reliable responses and a greater occurrence of unsupported legal interpretations. The integration of Retrieval-Augmented Generation (RAG), ChromaDB, and Guardrails significantly improved response quality by grounding outputs in official regulatory documents and reducing hallucinations.

The complete workflow can be reproduced using the referenced legal documents, PDF-to-Markdown conversion, translation through Gemini 2.5 Pro, ChromaDB indexing, and the Gemma 3 4B model deployed on infrastructure with specifications similar to those described above.

5. Results

RiskGovAI was evaluated across two functional modules. Given the hackathon timeframe, evaluation is qualitative and exploratory; no statistical significance testing was conducted.

Search Norms (Figure 1, see Appendix) supports natural-language queries filtered by jurisdiction. All five representative queries executed during the demonstration returned citation-backed responses referencing article-level passages from the indexed laws (Table 1), suggesting the Guardrails mechanism successfully constrained outputs to official sources. Citation presence, however, indicates grounding rather than legal accuracy; a formal precision/recall evaluation against expert-annotated answers would be required to substantiate reliability claims.

“**Figure 1** (see Appendix). Search Norms interface showing jurisdiction filters, a natural-language query field, and a query history panel with citation-backed responses. Available at: <https://riskgovai.com/>.”

Table 1. Five representative queries executed during the demonstration. All returned article-level citations from the indexed legal documents.

	Query topic	Jurisdictions	Citation returned
1	Social scoring AI restrictions	BR AI Act	Yes
2	Objection to automated performance evaluation	LFPDPPP	Yes
3	Transparency obligations for AI	BR AI Act	Yes
4	Data protection requirements	LGPD	Yes
5	Automated decision-making: Brazil $\times$ Chile	LGPD + Ley 21.719	Yes

Compare Jurisdictions (Figure 2, see Appendix) produces structured conflict reports. The illustrative scenario — automated credit decisions under LGPD and Ley 21.719 — returned 3 topics analyzed and 2 conflicts identified. This is directionally consistent with known differences between the two frameworks, but conflict classification was not validated against expert legal review.

“**Figure 2** (see Appendix). Compare Jurisdictions dashboard showing a saved comparison between LGPD and Ley 21.719, with 2 conflicts detected across 3 topics. Available at: <https://riskgovai.com/>”

New comparisons are created via a dedicated modal (Figure 3, see Appendix) requiring a minimum of two jurisdictions.

“**Figure 3** (see Appendix). *New Comparison modal with jurisdiction selectors (minimum 2 required).*”

Figure 4 (see Appendix) shows a side-by-side analysis of LGPD and Ley 21.719 across five regulatory aspects (automated decisions, consent, sensitive data, international transfer, and supervisory authority), each classified as Equivalent, Complementary, Partial, or Divergent. Observation: Four aspects were classified as Equivalent or Complementary; international transfer as Partial; and supervisory authority as Divergent, with article-level citations from both laws. Interpretation: The Divergent classification for supervisory authority reflects a known institutional difference between Brazil’s ANPD and Chile’s Data Protection

Agency. The Partial classification for international transfer appropriately flags the need for specialist review rather than resolving it automatically.

“**Figure 4** (see Appendix). Structured comparison between LGPD and Ley 21.719 across five regulatory aspects, with relation classification (Equivalent, Complementary, Partial, Divergent) and article-level citations from both jurisdictions.”

Figure 5 (see Appendix) presents the detailed topic-level view for the "Automated credit decisions" scenario, covering regulatory treatment, key regulatory differences, and compliance recommendation side by side for both jurisdictions. Observation: Both laws protect against fully automated decisions with review rights, but differ in explainability requirements. The system produced a unified compliance recommendation for both Brazilian and Chilean data subjects. Interpretation: This output demonstrates the platform's ability to synthesize actionable compliance guidance beyond simple document retrieval. It was not validated by legal professionals and should be treated as indicative rather than authoritative.

“**Figure 5** (see Appendix). Detailed topic-level comparison for the "Automated credit decisions" scenario, showing side-by-side regulatory treatment under LGPD and Ley 21.719, with conflict indicators and a unified compliance recommendation.”

The full stack ran on a 2 vCPU / 8 GB RAM VPS, confirming architectural viability under Global South infrastructure constraints.

6. Discussion and Limitations

Broader Implications for AI Safety

RiskGovAI's results suggest that RAG-grounded, scope-restricted legal AI can make regulatory compliance more accessible and auditable in the Global South. Organizations that cannot easily interpret data protection laws are more likely to deploy AI systems that violate rights or evade accountability; lowering the cost of regulatory research directly reduces this risk. The observed reduction in unsupported legal interpretations when using RAG+ Guardrails aligns with a broader AI safety principle: constraining outputs to verifiable sources is more reliable than parametric model knowledge alone in high-stakes domains.

Limitations

Evaluation was entirely qualitative; the five queries in Table 1 are not a representative sample, and no precision/recall benchmark was conducted. The central assumption, that citation presence indicates reliability, would not hold if the RAG pipeline retrieves plausible but contextually incorrect passages, producing cited yet legally wrong responses. The knowledge base covers only four of fourteen active data protection frameworks in Latin America, and AI-specific regulation (PL 2338/2023) is not indexed. English translation of source documents risks semantic drift on

terms without direct equivalents (e.g., habeas data). Conflict detection was not validated by legal professionals. Adversarial prompting, regulatory staleness, and sub-national regulations were not addressed. These constraints are partly attributable to the hackathon timeframe.

Future Work

Key next steps are: (1) quantitative benchmarking against expert-annotated answers; (2) knowledge base expansion to additional jurisdictions and AI-specific regulations; (3) multilingual retrieval to eliminate translation errors; (4) human-in-the-loop legal validation; (5) automated update mechanisms to keep indexed laws current; and (6) infrastructure upgrade to a higher-capacity VPS configuration (8 vCPU cores, 32 GB RAM, 400 GB NVMe storage, and 32 TB monthly bandwidth) at an estimated cost of US\$ 558.93/year, enabling support for larger models, faster response times, and concurrent user load as the platform scales beyond the current prototype.

6. Conclusion

RiskGovAI demonstrates that an open-source architecture combining Gemma 3 4B, RAG, ChromaDB, and Guardrails can automate regulatory research and cross-jurisdictional conflict detection for Latin American data protection law on infrastructure accessible to resource-constrained organizations. The prototype successfully delivered citation-backed regulatory search and structured comparative analysis across four jurisdictions (LGPD, Ley 21.719, Ley 1581, and LFPDPPP), running on a low-cost VPS at a fraction of the cost of proprietary legal AI solutions.

The broader implication for AI safety is that reliable, auditable legal AI does not require large budgets or specialized infrastructure. The RAG + Guardrails design, which constrains model outputs to verifiable official sources, offers a reproducible pattern for governance tools in regions historically underserved by existing legal tech solutions. Addressing the identified limitations, particularly quantitative evaluation and knowledge base expansion, would advance RiskGovAI from proof-of-concept toward a production-grade platform capable of meaningfully supporting responsible AI development across Latin America.

Code and Data

Code repository:

<https://github.com/elainefabiola/RiskGovAI---React-frontend-and-FastAPI-backend-using-RAG>

Data/Datasets:

<https://github.com/elainefabiola/RiskGovAI---React-frontend-and-FastAPI-backend-using-RAG/blob/main/backend/README.md>

Other artifacts <https://riskgovai.com/>

LLM Usage Statement

Claude (Anthropic) was used to support the drafting. Gemini 2.5 Pro was used for the translation of legal documents from Portuguese and Spanish into English before indexing. The platform interface screens were designed using Figma Maker.

References

- Brasil. (2018). **Lei nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD)**. Presidência da República. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm
- Chile. (2024). **Ley Nº 21.719 sobre Protección de Datos Personales**. Biblioteca del Congreso Nacional de Chile. Available at: <https://www.bcn.cl>
- Chroma. (2025). **ChromaDB Documentation**. Chroma. Available at: <https://docs.trychroma.com>
- Colombia. (2012). **Ley 1581 de 2012: Régimen General de Protección de Datos Personales**. Congreso de Colombia. Available at: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Google DeepMind. (2025). **Gemma 3 Technical Report**. Google DeepMind. Available at: <https://arxiv.org/abs/2503.19786>
- Google DeepMind. (2025). *Gemini 2.5 Technical Overview*. Google DeepMind. Available at: <https://deepmind.google/technologies/gemini/>
- Guha, N., Nyarko, J., Ho, D. E., Re, C., Cudaska, M., & LegalBench Authors. (2023). ***LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models***. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS 2023)*. <https://doi.org/10.48550/arXiv.2308.11462>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks**. *Advances in Neural Information Processing Systems (NeurIPS 2020)*. Available at: <https://arxiv.org/abs/2005.11401>
- México. (2010). **Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP)**. Cámara de Diputados del H. Congreso de la Unión. Available at: <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

OECD. (2019). **OECD Principles on Artificial Intelligence**. Organisation for Economic Co-operation and Development (OECD). Available at: <https://oecd.ai/en/ai-principles>

UNESCO. (2021). **Recommendation on the Ethics of Artificial Intelligence**. United Nations Educational, Scientific and Cultural Organization (UNESCO). Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models*. Available at: <https://arxiv.org/abs/2102.02503>

Appendix

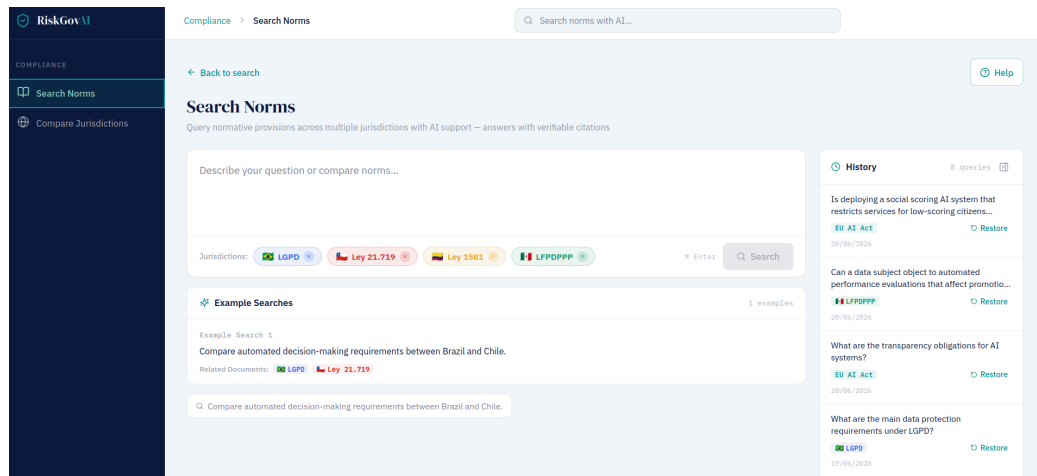


Figure 1 : Search Norms- <https://riskgovai.com/>

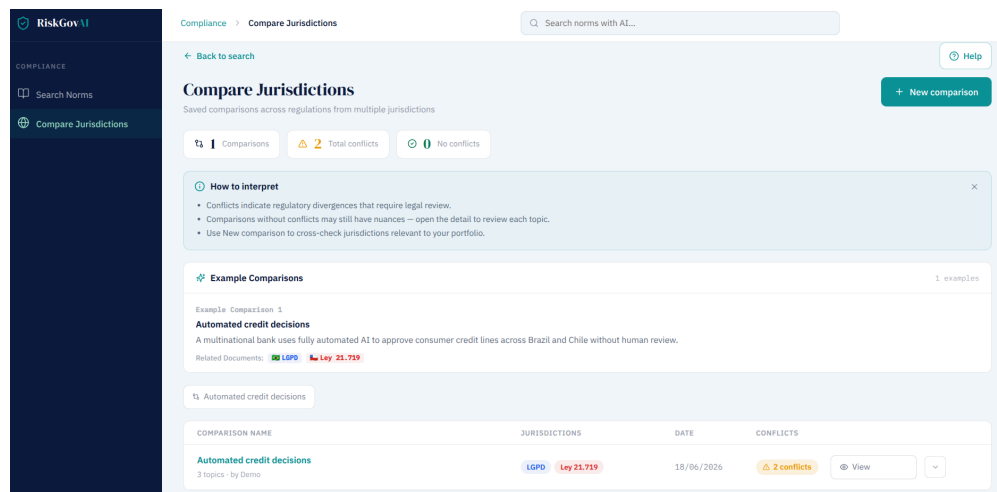


Figure 2: Compare Jurisdictions -<https://riskgovai.com/>

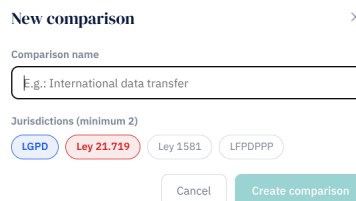


Figure 3: New comparison-<https://riskgovai.com/>

Q Compare automated decision-making requirements between Brazil and Chile. 🇧🇷 LGPD 🇨🇱 Ley 21.719 Q New query

How to interpret results

- Verified — excerpt validated in the indexed database; confirm applicability to your case.
- Partial confidence — likely relevance, but requires specialist review.
- Sources and reasoning — explains the normative cross-reference and possible conflicts.
- English excerpts display an option for automatic translation.

Structured comparison — 🇧🇷 LGPD 🇨🇱 Ley 21.719

Aspect	🇧🇷 LGPD	🇨🇱 Ley 21.719	Relation
Automated decisions	Art. 20 — right to review exclusively automated decisions (LGPD)	Art. 12 — right not to be subject to exclusively automated decisions (Ley 21.719)	Equivalent
Consent	Art. 7 — free, informed and unambiguous consent (LGPD)	Art. 4 — free, specific and informed consent (Ley 21.719)	Equivalent
Sensitive data	Art. 11 — processing subject to consent or legal grounds (LGPD)	Art. 9 — special categories with explicit consent (Ley 21.719)	Complementary
International transfer	Art. 33 — transfer subject to safeguards and ANPD authorization (LGPD)	Art. 28 — international transfer with adequate safeguards (Ley 21.719)	Partial
Supervisory authority	ANPD — National Data Protection Authority (Brazil)	Data Protection Agency — provided for in Ley 21.719 (Chile)	Divergent

The citations below complement this comparison with verifiable excerpts from each framework.

🇧🇷 LGPD — Brazil ⚠️ Partial confidence — review with specialist

“Article 63. The national authority shall establish rules on the progressive adjustment of databases created up to the date of entry into force of this Law, considering the complexity of the processing operations and the nature of the data. Article 64. The rights and principles expressed in this Law shall not exclude others provided for in the Brazilian legal system concerning the matter or in international treaties to which the Federative Republic of Brazil is a party.”

Art. 63 LGPD — Brazil Indexed RAG [View full text](#)

[↔️ Link to risk](#) [🔍 Add to comparison](#)

🇧🇷 LGPD — Brazil ⚠️ Partial confidence — review with specialist

“the Brazilian Internet Steering Committee; «Included by Law No. 13.853/2019» * «VII —> three (3) representatives from entities of the civil society with activities related to the protection of personal data; * «Included by Law No. 13.853/2019» * «VIII —> three (3) representatives from scientific, technological and innovative institutions; «Included by Law No. 13.853/2019» * «IX —> three (3) representatives from trade-union confederations representing the economic categories of the productive sector; «Included by Law No. 13.853/2019» * «X —> two (2) representatives from entities x”

Trecho LGPD — Brazil Indexed RAG [View full text](#)

[↔️ Link to risk](#) [🔍 Add to comparison](#)

Figure 4 - LGPD and Ley 21.719 - <https://riskgovai.com/>

RiskGov AI

Compliance > Compare Jurisdictions > Automated credit decisions Q Search norms with AI...

• Hover over the ? icons in the status legend for details.

Display: **All topics** Differences only Similarities only 3 of 3 topics

TOPIC	🇧🇷 LGPD	🇨🇱 Ley 21.719
Regulatory treatment ⚠️ Conflict	LGPD Article 20 grants data subjects the right to request review of decisions made solely on automated processing affecting their interests, including credit profiling. Controllers must provide clear information on criteria and procedures used, subject to trade secrets. Article 20 grants data subjects the right to request review of decisions made solely on automated processing affecting their interests · LGPD ⚠️ Conflict	Law 21.719 Article 9 bis grants the right not to be subject to automated decisions including profiling that produce legal effects or significantly affect the data subject. Controllers must ensure transparency, explanation, human intervention, and the right to request decision review in all automated decision cases. Article 9 bis grants the right not to be subject to automated decisions including profiling that produce legal effects or significantly affect the data subject · Ley 21.719 ⚠️ Conflict
Key regulatory difference ⚠️ Conflict	Both jurisdictions protect against solely automated decisions with review rights, but Chile explicitly mandates human intervention, explanation, and transparency measures in all automated decision scenarios (including exceptions), while LGPD focuses on the right to request review and information on criteria. — ⚠️ Conflict	Both jurisdictions protect against solely automated decisions with review rights, but Chile explicitly mandates human intervention, explanation, and transparency measures in all automated decision scenarios (including exceptions), while LGPD focuses on the right to request review and information on criteria. — ⚠️ Conflict
Compliance recommendation	Implement a human review workflow and explainability layer for credit decisions in both jurisdictions. Document decision logic and provide a unified contestation channel accessible to Brazilian and Chilean data subjects. — 🟢 Allowed	Implement a human review workflow and explainability layer for credit decisions in both jurisdictions. Document decision logic and provide a unified contestation channel accessible to Brazilian and Chilean data subjects. — 🟢 Allowed

Status: ✔️ Required 🟢 Equivalent 🟢 Allowed ⚪ Not addressed ⚠️ Conflict

Figure 5 - Automated credit decisions - <https://riskgovai.com/>