

Comunicação

Soberania e Governança de Dados de Povos Indígenas em Tempos de Inteligência Artificial Generativa

*Data Sovereignty and Governance of Indigenous Peoples in the
Era of Generative Artificial Intelligence*

*Soberanía y Gobernanza de Datos de los Pueblos Indígenas en la
Era de Inteligencia Artificial Generativa*

Laís Barbudo Carrasco*

Doutora em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil

Professora Doutora do Programa de Pós-Graduação em Ciência da Informação e Arquivologia, Mid Sweden University (MIUN), Sundsvall, Västernorrland, Suécia

ORCID: <https://orcid.org/0000-0003-2938-9390>

Lattes: <http://lattes.cnpq.br/7793712097080187>

E-mail: laís.barbudo@miun.se

Emanuelle Torino

Doutora em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil

Bibliotecária da Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, Paraná, Brasil
Coordenadora da Coordenação de Serviços Bibliográficos (COBIB), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, Distrito Federal, Brasil

ORCID: <https://orcid.org/0000-0002-3791-9884>

Lattes: <http://lattes.cnpq.br/5042296869081637>

E-mail: emanuelle@utfpr.edu.br

Caio Saraiva Coneglan

Doutor em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil

Professor e coordenador dos cursos de graduação de Análise e Desenvolvimento de Sistemas e Ciência da Computação da Universidade de Marília (Unimar), Marília, São Paulo, Brasil

Professor do Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil
ORCID: <https://orcid.org/0000-0002-6126-9113>
Lattes: <http://lattes.cnpq.br/3954065076810604>
E-mail: caio.coneglian@gmail.com

Silvana Aparecida Borsetti Gregorio Vidotti

Doutora em Educação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil

Professora do Departamento e do Programa de Pós-Graduação em Ciência da Informação,

Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil

Coordenadora de Tecnologias Aplicadas (COTEA), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, Distrito Federal, Brasil

ORCID: <https://orcid.org/0000-0002-4216-0374>

Lattes: <http://lattes.cnpq.br/7390573927636069>

E-mail: silvana.vidotti@unesp.br

Resumo

O trabalho discute os desafios e as oportunidades que a Inteligência Artificial Generativa (IAG) impõe à soberania e à governança de dados dos povos indígenas. A crescente utilização de grandes volumes de dados para o treinamento de algoritmos de IAG, muitas vezes sem consentimento ou reconhecimento das fontes, levanta sérias questões éticas relacionadas à propriedade intelectual, à representação cultural e à autodeterminação. Nesse contexto, os princípios CARE (Benefício Coletivo, Autoridade para Controlar, Responsabilidade e Ética) emergem como diretrizes fundamentais para garantir o uso ético dos dados indígenas, complementando os princípios FAIR, que tratam da interoperabilidade técnica dos dados. O estudo também destaca a importância das Local Contexts Labels (Etiquetas de Contextos Locais), como a Traditional Knowledge (TK) e a Biocultural (BC), que permitem às comunidades indígenas sinalizar condições específicas de acesso, uso e circulação de seus dados e conhecimentos tradicionais. Essas ferramentas fortalecem a governança cultural e a proteção contra usos indevidos por sistemas de IAG. A pesquisa, de base teórico-exploratória, defende a incorporação dessas práticas nas políticas de desenvolvimento de IAG. Conclui-se que a inclusão efetiva das comunidades indígenas nos processos de governança de dados e no desenvolvimento tecnológico é essencial para garantir justiça social e respeito à soberania informacional indígena.

Palavras-chave: Governança e soberania de dados indígenas; Inteligência Artificial Generativa (IAG); Princípios CARE; Etiquetas Local Contexts.

Abstract

This paper examines the challenges and opportunities that Generative Artificial Intelligence (GAI) poses to the data sovereignty and governance of Indigenous Peoples. The increasing use of large datasets to train GAI algorithms — often without consent or acknowledgment of the sources — raises serious ethical concerns related to intellectual property, cultural representation, and self-determination. In this context, the CARE Principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) emerge as essential

guidelines to ensure the ethical use of Indigenous data, complementing the FAIR Principles, which address the technical interoperability of data. The study also underscores the importance of Local Contexts Labels, such as Traditional Knowledge (TK) and Biocultural (BC) labels, which enable Indigenous communities to establish specific conditions for access, use, and circulation of their data and traditional knowledge. These tools strengthen cultural governance and provide safeguards against misuse of GAI systems. Adopting a theoretical and exploratory approach, the research advocates for the integration of these practices into GAI development policies. It concludes that the meaningful inclusion of Indigenous communities in data governance and technological development processes is essential to ensure social justice and respect for Indigenous informational sovereignty.

Keywords: Indigenous Data Governance and Sovereignty; Generative Artificial Intelligence (GAI); CARE Principles; Local Contexts Labels.

Resumen

Este trabajo analiza los desafíos y las oportunidades que la Inteligencia Artificial Generativa (IAG) plantea para la soberanía y la gobernanza de los datos de los Pueblos Indígenas. El uso creciente de grandes volúmenes de datos para el entrenamiento de algoritmos de IAG —a menudo sin consentimiento ni reconocimiento de las fuentes— plantea serias cuestiones éticas relacionadas con la propiedad intelectual, la representación cultural y la autodeterminación. En este contexto, los Principios CARE (Beneficio Colectivo, Autoridad para Controlar, Responsabilidad y Ética) emergen como directrices fundamentales para garantizar el uso ético de los datos indígenas, complementando los Principios FAIR, que abordan la interoperabilidad técnica de los datos. El estudio también destaca la importancia de las Etiquetas de Contextos Locales (Local Contexts Labels), como las de Conocimiento Tradicional (TK) y Biocultural (BC), que permiten a las comunidades indígenas señalar condiciones específicas de acceso, uso y circulación de sus datos y conocimientos tradicionales. Estas herramientas fortalecen la gobernanza cultural y ofrecen protección contra el uso indebido por parte de sistemas de IAG. Basada en un enfoque teórico-exploratorio, la investigación defiende la incorporación de estas prácticas en las políticas de desarrollo de la IAG. Se concluye que la inclusión efectiva de las comunidades indígenas en los procesos de gobernanza de datos y desarrollo tecnológico es esencial para garantizar la justicia social y el respeto a la soberanía informacional indígena.

Palabras clave: Gobernanza y Soberanía de Datos Indígenas; Inteligencia Artificial Generativa (IAG); Principios CARE; Etiquetas de Contextos Locales.

Introdução

A ampliação da capacidade das máquinas de consumir dados em larga escala aliada à redução dos custos de armazenamento resultou em um cenário de redundância de dados em diferentes esferas. Tal contexto apresenta benefícios ao possibilitar que as infraestruturas permaneçam operacionais, contribuindo inclusive para aspectos relacionados à preservação digital. Contudo, é necessário atentar para os riscos decorrentes desse processo, especialmente no que se refere à segurança da informação e à proteção de dados pessoais e sensíveis. Além disso, a abertura de dados busca alternativas técnicas que otimizem o fornecimento e o consumo de dados, visando assegurar que sejam automatizados, de modo que o processamento e o reuso apresentem custos inferiores aos benefícios gerados.

O cenário torna-se ainda mais desafiador com o desenvolvimento e a massificação da Inteligência Artificial Generativa (IAG), que passou a consumir e utilizar quantidades imensuráveis de dados disponíveis em ambientes digitais para que tais ferramentas apresentem um comportamento adequado. Tal utilização dos dados para o treinamento da Inteligência Artificial acontece de forma automática e abrangente, a partir da obtenção de dados provenientes dos mais diversos ambientes informacionais digitais. Dessa forma, não há um controle efetivo sobre quais dados foram utilizados, o que dificulta a definição de aspectos éticos e legais, como, por exemplo, a propriedade intelectual.

Para enfrentar esses desafios, são necessários tanto recursos técnicos quanto diretrizes éticas. Os princípios FAIR – Findable, Accessible, Interoperable e Reusable – visam garantir que os dados possam ser localizados e reutilizados eficientemente por máquinas. No entanto, esses princípios não contemplam aspectos éticos e socioculturais, o que os torna limitados diante de dados sensíveis, como os pertencentes aos povos indígenas. Como resposta, foram elaborados os princípios CARE – Collective Benefit, Authority to Control, Responsibility e Ethics –, que colocam em foco a autodeterminação, a governança cultural e os direitos coletivos dos povos indígenas sobre seus dados (Carroll et al., 2020). Assim, os princípios CARE complementam os princípios FAIR ao integrar princípios éticos às práticas técnicas.

A iniciativa Local Contexts contribui nesse sentido, ao oferecer etiquetas (*labels*) que incorporam metadados jurídicos e culturais, permitindo que as comunidades indígenas definam os termos de uso, acesso e interpretação de seus conhecimentos tradicionais e bioculturais (Local Contexts, 2025a; Christen, 2015). O uso das etiquetas Local Contexts na representação de dados de povos indígenas dialoga com o disposto no Artigo 31 da Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas:

1. Os povos indígenas têm o direito de manter, controlar, proteger e desenvolver seu patrimônio cultural, seus conhecimentos tradicionais, suas expressões culturais tradicionais e as manifestações de suas ciências, tecnologias e culturas, compreendidos os recursos humanos e genéticos, as sementes, os medicamentos, o conhecimento das propriedades da fauna e da flora, as tradições orais, as literaturas, os desenhos, os esportes e jogos tradicionais e as artes visuais e interpretativas. Também têm o direito de manter, controlar, proteger e desenvolver sua propriedade intelectual sobre o mencionado patrimônio cultural, seus conhecimentos tradicionais e suas expressões culturais tradicionais.
2. Em conjunto com os povos indígenas, os Estados adotarão medidas eficazes para reconhecer e proteger o exercício desses direitos. (Nações Unidas, 2008, p. 16).

Nesse contexto, a soberania e a proteção dos dados de povos indígenas podem entrar em conflito com as ferramentas de Inteligência Artificial Generativa, caso preceitos éticos e legais sejam desconsiderados. Esse fato reforça a necessidade de compreender que a representação da informação é imprescindível para o fornecimento e o consumo

de dados e, consequentemente, para que o uso desses dados por ferramentas de IA possa ser realizado de forma adequada.

Diante do exposto, este estudo propõe-se a abordar elementos técnicos e éticos que possibilitem a interoperabilidade dos dados entre sistemas, inclusive os de Inteligência Artificial, considerando a autonomia, a soberania e a governança de dados pelos povos indígenas. Trata-se de um estudo de base teórico-exploratória, realizado a partir de levantamento bibliográfico sobre as temáticas abordadas, visando embasar a discussão apresentada.

Inteligência Artificial Generativa e a Soberania e Governança de Dados dos Povos Indígenas

Dados indígenas referem-se a informações relacionadas aos povos indígenas, seus sistemas de conhecimento, suas terras, comunidades e culturas. Esses dados podem ser tanto quantitativos, como estatísticas demográficas e de saúde, quanto qualitativos, abrangendo narrativas culturais, conhecimentos tradicionais e outras formas de propriedade intelectual indígena (Kukutai; Taylor, 2016). Dados indígenas são distintos de outras formas de dados porque estão intrinsecamente ligados às identidades, à soberania e à autodeterminação dos povos indígenas. Essa conexão impõe uma responsabilidade ética e de governança significativa acerca de como tais dados são coletados, gerenciados e utilizados.

A soberania de dados indígenas (SDI) refere-se ao direito dos povos indígenas de governar a coleta, a propriedade e a aplicação de dados sobre si mesmos, suas terras e suas comunidades. Está enraizada no conceito mais amplo de soberania indígena, que afirma a autoridade inerente das comunidades indígenas de se autogovernarem sem interferência externa (Rainie et al., 2019). A SDI é um aspecto crítico da luta mais ampla pelos direitos e autodeterminação indígenas, pois assegura que os povos indígenas mantenham o controle sobre seus dados, o que é crucial para preservar sua integridade cultural e proteger sua propriedade intelectual (Walter et al., 2021).

Kukutai e Taylor (2016) argumentam que as práticas atuais de coleta e gestão de dados muitas vezes falham em reconhecer os direitos soberanos dos povos indígenas, levando ao uso de dados de maneiras que não estão alinhadas aos seus valores ou interesses. Isso pode resultar em má representação, exploração e danos às comunidades indígenas. Portanto, o movimento em direção à soberania de dados de povos indígenas é uma resposta a esses desafios, ao defender o desenvolvimento de políticas e estruturas que respeitem os direitos indígenas e priorizem sua autodeterminação na governança de dados.

As estruturas de governança desempenham um papel crucial em garantir que a soberania de dados indígenas seja respeitada e mantida. Boas práticas de governança incluem o estabelecimento de políticas e diretrizes claras que ditem como os dados de povos indígenas devem ser coletados, armazenados e utilizados. Essas práticas visam

proteger a privacidade, a integridade e a propriedade dos dados indígenas, assegurando que sejam utilizados de maneiras que beneficiem as comunidades indígenas (Lovett et al., 2019).

Snipp (2016) destaca que a governança de dados de povos indígenas envolve não apenas a gestão de dados, mas também a criação de um ambiente no qual os povos indígenas possam exercer sua soberania em questões relacionadas aos dados. Isso inclui o desenvolvimento de infraestruturas de dados controladas por comunidades indígenas, bem como a criação de estruturas legais que protejam os dados indígenas contra o uso indevido por entidades externas.

Na prática, uma governança eficaz de dados de povos indígenas requer colaboração entre comunidades indígenas, pesquisadores e formuladores de políticas. Harding et al. (2012) discutem a importância de estabelecer parcerias baseadas em confiança, respeito mútuo e objetivos compartilhados. Essas parcerias são essenciais para assegurar que as práticas de governança de dados sejam culturalmente apropriadas e alinhadas com as necessidades e prioridades das comunidades indígenas.

Apesar dos avanços na promoção da soberania de dados de povos indígenas, ainda existem desafios significativos. Entre eles, estão a necessidade de maior reconhecimento dos direitos indígenas nos marcos legais nacionais e internacionais, bem como o desenvolvimento de capacidades técnicas e institucionais dentro das comunidades indígenas para gerenciar e governar seus dados de forma eficaz (Kukutai; Taylor, 2016). Além disso, é necessário que haja diálogo contínuo e colaboração entre as comunidades indígenas e outros stakeholders para garantir que as práticas de governança de dados continuem a evoluir de maneiras que respeitem a soberania indígena e promovam o bem-estar dessas comunidades (Walter; Suina, 2023).

A governança dos dados indígenas é uma questão crítica que cruza com preocupações mais amplas sobre a soberania e autodeterminação indígenas. À medida que o movimento pela soberania de dados indígenas continua a crescer, é essencial que as estruturas de governança sejam desenvolvidas e implementadas de maneiras que respeitem os direitos indígenas e empoderem as comunidades indígenas. Isso requer um compromisso com boas práticas de governança, colaboração contínua e o reconhecimento dos sistemas de conhecimento indígenas como parte integrante da gestão e do uso dos dados indígenas. Sendo assim, destaca-se a necessidade de defesa e ação contínuas para assegurar que os dados de povos indígenas sejam governados de forma alinhada com os valores e as aspirações desses povos.

Em tempos de Inteligência Artificial Generativa (IAG), a questão da soberania e governança de dados de povos indígenas torna-se ainda mais complexa e urgente. A tecnologia de IAG, como modelos de linguagem e de geração de imagens, utiliza grandes quantidades de dados para treinar algoritmos capazes de criar novos conteúdos a partir de padrões aprendidos.

Vale destacar como a IAG é definida:

Com capacidades para gerar conteúdo contextualmente relevante e de alta qualidade, quase indistinguível do trabalho criado por humanos [...]. A IA generativa encontra sua utilidade em diversas modalidades, incluindo a geração de texto, imagem, vídeo, código, som e outros conteúdos produzidos, como moléculas ou renderizações 3D. (Banh; Strobel, 2023, tradução nossa).

Dentro desse cenário, a IAG tem sido utilizada em vários contextos, especialmente na área de geração de textos e imagens. No entanto, uma das principais características da IAG está na utilização de grandes quantidades de dados para o treinamento dessas ferramentas. Com isso, tem-se uma discussão acerca da propriedade intelectual e dos direitos dos dados que foram utilizados.

Embora essas inovações ofereçam oportunidades significativas em áreas como educação, saúde e preservação cultural, elas também apresentam desafios éticos e práticos que podem ameaçar os direitos dos povos indígenas, especialmente no que diz respeito ao controle de seus dados e conhecimentos tradicionais.

Um dos principais desafios que a IAG impõe à soberania de dados indígenas está na coleta e no uso de dados em larga escala. Muitas vezes, os dados utilizados para treinar algoritmos de IA são extraídos da internet sem o devido consentimento ou reconhecimento das fontes originais, o que pode incluir informações culturais, histórias e conhecimentos tradicionais indígenas. Esse processo de coleta e reutilização de dados levanta questões sobre consentimento informado, propriedade intelectual e exploração cultural.

Além disso, os algoritmos de IAG podem reproduzir estereótipos ou deturpar culturas indígenas ao gerar novos conteúdos baseados em padrões de dados incompletos ou incorretos. Como esses sistemas são projetados para replicar e amplificar os dados nos quais são treinados, há um risco real de que narrativas indígenas sejam distorcidas ou apropriadas, perpetuando preconceitos e contribuindo para a desinformação.

Embora a IAG apresente riscos, também oferece oportunidades promissoras para a preservação cultural e o fortalecimento da soberania de dados indígenas. Tecnologias de IAG podem ser usadas para ajudar na documentação e transmissão de histórias, práticas culturais e línguas indígenas, desde que essas iniciativas sejam lideradas pelas próprias comunidades indígenas. Por exemplo, a geração automática de textos e imagens pode ser utilizada para criar materiais educativos em línguas indígenas ou para simular ambientes culturais e históricos que podem apoiar o ensino e a valorização das tradições indígenas.

No entanto, para que essas tecnologias realmente beneficiem as comunidades indígenas, elas devem ser implementadas em um contexto de soberania digital. Isso significa que as comunidades indígenas devem ter controle total sobre como suas

culturas e conhecimentos são representados e utilizados por sistemas de IA. Além disso, os desenvolvedores de IA devem adotar uma abordagem de desenvolvimento que incorpore as perspectivas e os conhecimentos indígenas desde o início do processo.

Para proteger a soberania de dados de povos indígenas em tempos de Inteligência Artificial Generativa, é essencial que as estruturas de governança de dados se adaptem às novas realidades tecnológicas. Isso envolve a criação de diretrizes específicas para o uso de dados de povos indígenas em tecnologias de IA, assegurando que eles sejam utilizados de maneira que respeite os direitos, os interesses, as visões de mundo e a autodeterminação das comunidades indígenas.

Princípios CARE

A coleta, o gerenciamento e a utilização de dados indígenas têm sido historicamente moldados por entidades externas, muitas vezes resultando na marginalização das perspectivas e interesses indígenas. Consequentemente, há um reconhecimento crescente da necessidade de estruturas que assegurem a governança ética dos dados indígenas, enfatizando a soberania e autodeterminação das comunidades indígenas (Carroll et al., 2020).

Em resposta à crescente demanda por uma abordagem mais ética à governança de dados de povos indígenas, foram desenvolvidos os princípios CARE, um acrônimo para Collective Benefit (Benefício Coletivo), Authority to Control (Autoridade para Controlar), Responsibility (Responsabilidade) e Ethics (Ética). Os princípios CARE foram estabelecidos como uma estrutura complementar aos já existentes princípios FAIR, que estabelecem subprincípios para favorecer a localização, o acesso, a interoperabilidade e o reuso de dados. No entanto, ao contrário dos princípios FAIR, que focam principalmente nos aspectos técnicos necessários ao processamento automático de dados, os princípios CARE enfatizam o tratamento ético e respeitoso dos dados de povos indígenas, considerando seus direitos e suas necessidades (Carroll et al., 2020; Vidotti; Torino; Coneglian, 2021; Torino; Monteiro; Vidotti, 2024).

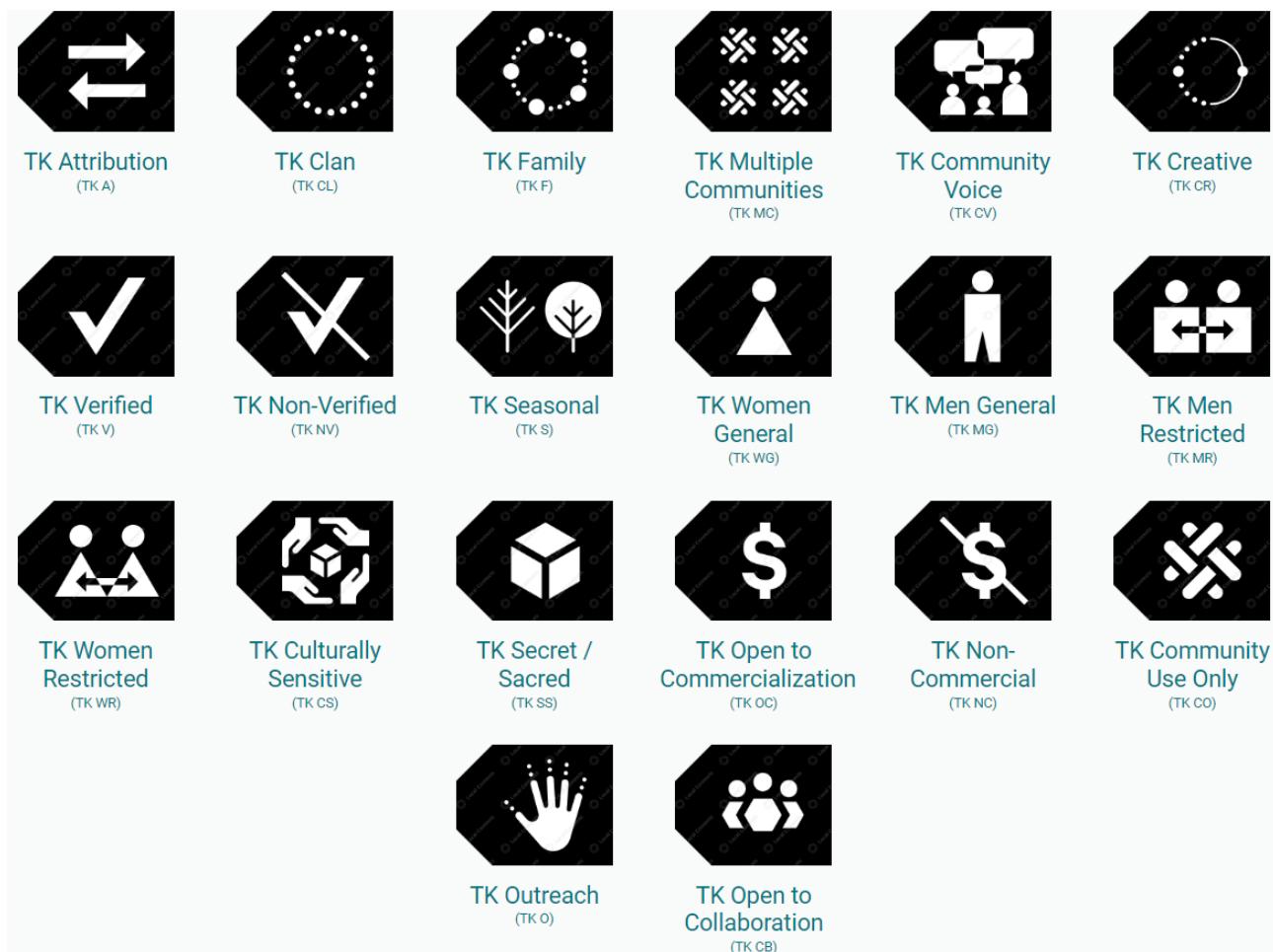
Os princípios CARE foram articulados em 2019 pela Global Indigenous Data Alliance (GIDA) como uma resposta às inadequações das estruturas de governança de dados existentes que frequentemente ignoram as perspectivas indígenas. Com isso, os princípios CARE visam abordar as injustiças históricas enfrentadas pelas comunidades indígenas em relação à propriedade e ao uso de seus dados, defendendo um modelo de governança de dados que seja ao mesmo tempo ético e equitativo (Carrasco, Vidotti, Torino, 2024).

A operacionalização dos princípios CARE envolve incorporá-los nas políticas e práticas de gestão de dados, visando assegurar que os dados de povos indígenas sejam governados de uma forma alinhada aos seus valores e interesses.

Local Contexts Labels

A iniciativa Local Contexts oferece um conjunto de etiquetas (*labels*) que podem ser aplicadas a registros digitais, definindo o contexto cultural e legal de seu uso. As Etiquetas de Conhecimento Tradicional (TK) e Etiquetas Bioculturais (BC) refletem os valores e protocolos das comunidades indígenas, promovendo autodeterminação informacional e governança cultural (Local Contexts, 2025b).

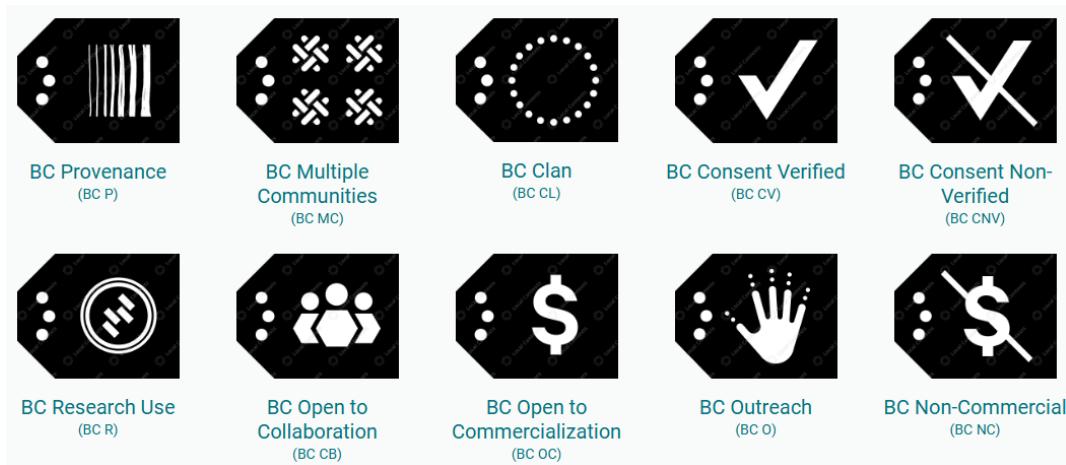
Figura 1. Identidade visual das etiquetas TK



Fonte: *TK Labels – Local Contexts* (2025c)

As TK Labels incluem categorias como TK Attribution, que corrige atribuições históricas incorretas; TK Community Voice, que reconhece autoria coletiva; e TK Culturally Sensitive, que indica restrições relacionadas a gênero, sazonalidade ou aspectos sagrados. Já as Etiquetas de Permissão como TK Community Use Only e TK Open to Collaboration estabelecem condições para o uso e compartilhamento dos objetos.

Figura 2. Identidade visual das etiquetas BC



Fonte: *BC Labels – Local Contexts* (2025d)

As BC Labels, por sua vez, ampliam esse sistema para dados genéticos e coleções biológicas. A BC Provenance identifica o interesse indígena sobre dados científicos, enquanto a BC Consent Verified assegura que o consentimento foi obtido corretamente. Essas etiquetas funcionam como marcadores de proveniência, expectativas culturais e limitações de uso, essenciais em contextos digitais e automatizados (Local Contexts, 2025d). Essas ferramentas permitem às comunidades indígenas controlar e contextualizar o uso de seus dados, prevenindo a apropriação indevida e promovendo a justiça informacional.

Inteligência Artificial Generativa e a Soberania dos Dados Indígenas

O processo de treinamento de soluções e algoritmos de Inteligência Artificial Generativa requer a utilização de quantidades imensuráveis de textos e conteúdos. Nesse sentido, por exemplo, uma solução que gere textos de forma automática (como o ChatGPT) necessita obter e processar milhões de páginas da web e conteúdos disponíveis de forma aberta na internet.

Além disso, quanto maior for a diversidade de conteúdos obtidos, melhores serão as respostas e o espectro de atuação da ferramenta. Contudo, não existe uma gestão e um controle efetivo acerca da propriedade intelectual e de outros aspectos éticos, como elementos discriminatórios que possam estar presentes nos dados que foram utilizados para o treinamento das ferramentas.

Ao refletir sobre o objetivo e o esforço para que os povos indígenas tenham soberania acerca dos seus dados, nota-se que há um evidente conflito em como as tecnologias de Inteligência Artificial Generativa estão utilizando os mais diversos documentos para o treinamento dos seus algoritmos. Isso ocorre porque não há, atualmente, um meio de identificar e validar quais foram os documentos utilizados para o treinamento das ferramentas, como ChatGPT ou Gemini do Google, nem as características dos documentos em questão.

A partir desse cenário, o advento da Inteligência Artificial Generativa apresenta um desafio significativo, com o potencial de comprometer os esforços que as comunidades indígenas vêm realizando ao longo dos anos para assegurar a soberania sobre seus dados. A falta de transparência em relação aos documentos utilizados para treinar essas ferramentas aliada à ausência de controle por parte dessas comunidades pode resultar na exploração indevida de seus dados para alimentar tecnologias desenvolvidas por grandes corporações, muitas delas pertencentes a potências ocidentais. Isso levanta preocupações éticas profundas, pois perpetua a assimetria de poder e ameaça o direito dos povos indígenas de decidir como seus conhecimentos e patrimônios culturais são representados e utilizados no ambiente digital.

O cenário apresentado traz a necessidade de adotar padrões e tecnologias que sinalizem e apontem que determinados dados não podem ser utilizados pelas ferramentas da forma que está sendo realizada atualmente. Para tal, os princípios CARE e as etiquetas Local Contexts podem servir de apoio ao processo de sinalização de que determinados dados não podem ser utilizados pelas ferramentas de Inteligência Artificial ou indicar o modo como tais dados podem ser utilizados.

Evidentemente, para isso, há a necessidade de que as ferramentas de grandes empresas de Inteligência Artificial Generativa, como Google, OpenAI, Amazon e Microsoft, se comprometam a identificar a representação dos documentos e considerem os aspectos representados pelas etiquetas e pelos princípios CARE no processo da coleta e do processamento dos dados.

Diante disso, foram definidas algumas etiquetas que, caso consideradas pelas ferramentas de Inteligência Artificial Generativa, podem apoiar na manutenção da soberania dos dados dos povos indígenas. O Quadro 1 mostra algumas ações que as ferramentas de IA Generativa podem ter na utilização dos dados de povos indígenas.

Quadro 1. Ações das ferramentas para manutenção da soberania dos dados

Ação	Descrição
Apresentação das etiquetas <i>Local Contexts</i>	Ao gerar conteúdo que foi criado a partir dos dados que são propriedade ou vinculados aos povos indígenas, as ferramentas de Inteligência Artificial Generativa devem apresentar as etiquetas descrevendo aspectos relacionados à cultura e à propriedade dos dados.
Evitar a coleta de dados que forem sinalizados	Há algumas etiquetas segundo as quais os dados não podem ser utilizados por outras culturas e ferramentas. Dessa forma, no processo de coleta dos dados para a utilização em ferramentas de IA Generativa, deve ser evitada a coleta de tais dados, o que impediria que os dados dos povos indígenas fossem utilizados de forma indevida.
Não utilizar conteúdos definidos como não comerciais	Outro contexto que deve evitar a coleta dos dados diz respeito a etiquetas que proíbem a utilização de dados para fins comerciais. Como as ferramentas de IA Generativa podem ser utilizadas nos mais diversos contextos, inclusive para fins comerciais, os dados etiquetados como não comerciais não podem ser coletados pelas ferramentas.

Atender aos princípios CARE durante a coleta e utilização dos dados no treinamento	De uma forma mais ampla, realiza-se a proposta de utilizar os dados de povos indígenas observando os princípios CARE (Benefício Coletivo, Autoridade para Controlar, Responsabilidade e Ética), considerando que os dados dessas comunidades devem ser respeitados. Além disso, é necessário que as ferramentas tenham a compreensão de que os povos indígenas têm soberania sobre os seus dados e, portanto, esses povos precisam ser chamados para a discussão acerca do que fazer com os seus dados e de como fazê-lo.
------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fonte: elaborado pelos autores.

A partir do que foi apresentado no Quadro 1, verifica-se que há um contexto ético para as ferramentas, pois grande parte das ações exige que elas atendam a esses critérios, sem a capacidade de verificar isso de uma forma clara. Dessa forma, é necessário que a discussão sobre os contextos éticos da Inteligência Artificial Generativa inclua a proteção à soberania dos dados dos povos indígenas em sua agenda.

Considerações Finais

A Inteligência Artificial Generativa representa tanto uma oportunidade quanto uma ameaça à soberania dos dados indígenas. Se, por um lado, ela pode apoiar iniciativas culturais e educacionais, por outro, pode perpetuar assimetrias de poder, exploração cultural e invisibilização de direitos coletivos.

Neste estudo, argumenta-se que a adoção dos princípios CARE e das etiquetas Local Contexts constitui um caminho promissor para mitigar os impactos negativos da IAG e promover uma governança mais justa e ética dos dados indígenas. As etiquetas oferecem às comunidades um mecanismo direto de controle e representação de seus dados, enquanto os princípios CARE fornecem a base ética necessária para decisões sobre coleta, uso e reuso de informações.

Entretanto, a eficácia dessas iniciativas depende do engajamento ativo das grandes corporações tecnológicas e da inclusão da autodeterminação indígena como valor central nas agendas de desenvolvimento tecnológico. A justiça algorítmica e a soberania informacional dos povos indígenas não podem ser tratadas como apêndices, mas sim como componentes centrais de uma IA verdadeiramente ética e inclusiva.

Para que a inovação tecnológica seja compatível com os direitos humanos e a justiça social, é indispensável o compromisso com padrões que respeitem as comunidades mais vulneráveis. Somente por meio da colaboração, da transparência e do respeito à diversidade epistêmica será possível construir um futuro digital mais equitativo.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Não aplicável.

CRediT – Contribuições dos Autores

Laís Barbudo Carrasco | Concretualização, Escrita – redação original, Supervisão
Emanuelle Torino | Concretualização, Escrita – redação original
Caio Saraiva Coneglian | Escrita – revisão e edição
Silvana Aparecida Borsetti Gregorio Vidotti | Escrita – revisão e edição

Referências

- Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 63. <https://doi.org/10.1007/s12525-023-00680-1>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., ... & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
- Carrasco, L. B., Vidotti, S. A. B. G., & Torino, E. (2024, junho). Arquivos, princípios CARE e a governança de dados indígenas. In VII Workshop de Informação, Dados e Tecnologia – WIDaT 2024 (Vol. 7). <https://doi.org/10.22477/vii.widat.194>
- Christen, K. (2015). Tribal archives, traditional knowledge, and local contexts: Why the “s” matters. *Journal of Western Archives*, 6(1), Article 3. <https://doi.org/10.26077/78d5-47cf>
- GIDA – Global Indigenous Data Alliance. (2019). CARE Principles for Indigenous Data Governance. CARE Principles for Indigenous Data Governance_FINAL_Sept 06 2019. pdf
- Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., & Donatuto, J. (2012). Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environmental Health Perspectives*, 120(1), 6–10. <https://doi.org/10.1289/ehp.1103904>
- Kukutai, T., & Taylor, J. (2016). Data sovereignty for indigenous peoples: Current practice and future needs. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 39–55). Canberra: ANU Press.
- Local Contexts. (2024a). Grounding Indigenous rights. <https://localcontexts.org/>
- Local Contexts. (2024b). About Labels. <https://localcontexts.org/labels/about-the-labels/>
- Local Contexts. (2024c). TK Labels. <https://localcontexts.org/labels/traditional-knowledge-labels/>
- Local Contexts. (2024d). BC Labels. <https://localcontexts.org/labels/biocultural-labels/>
- Lovett, R., Lee, V., Kukutai, T., Cormack, D., & Rainie, S. C. (2019). Good data practices for Indigenous data sovereignty and governance. In C. Dalton & L. Taylor (Eds.), *Good data* (pp. 26–36). Amsterdam: Institute of Network Cultures.
- Nações Unidas. (2008). Declaração das Nações Unidas sobre os direitos dos povos

indígenas. https://www.un.org/esa/socdev/unpfii/documents/DRIPS_pt.pdf

Rainie, S. C., Kukutai, T., Walter, M., Figueroa-Rodríguez, O. L., Walker, J., & Axelsson, P. (2019). Indigenous data sovereignty.

Snipp, C. M. (2016). What does data sovereignty imply: What does it look like. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 39–55). ANU Press.

Torino, E., Carrasco, L. B., & Vidotti, S. A. B. G. (2024). Representação de objetos indígenas em ambientes digitais: A iniciativa Local Contexts. In VII Workshop de Informação, Dados e Tecnologia – WIDaT 2024 (Vol. 7). <https://doi.org/10.22477/vii.widat.195>

Torino, E., Monteiro, E. C. S. A., & Vidotti, S. A. B. G. (2023). Plano de gestão de dados de pesquisa de povos indígenas: Considerações acerca dos princípios FAIR e CARE. *Revista Brasileira de Preservação Digital*, 4.

Vidotti, S. A. B. G., Torino, E., & Coneglian, C. S. (2021). #SejaJUSTOeCUIDADOSO: Princípios FAIR e CARE na gestão de dados de pesquisa. In L. F. Sales et al. (Eds.), *Princípios FAIR aplicados à gestão de dados de pesquisa* (pp. 209–221). Rio de Janeiro: Ibict.

Vidotti, S. A. B. G., Torino, E., & Coneglian, C. S. (2021). #SejaJUSTOeCUIDADOSO: Princípios FAIR e CARE na gestão de dados de pesquisa. In L. F. Sales, V. S. Veiga, P. Henning, & L. F. Sayão (Orgs.), *Princípios FAIR aplicados à gestão de dados de pesquisa* (pp. 201–214). Ibict. <https://doi.org/10.22477/9786589167242.cap15>

Walter, M., Suina, M., Lovett, R., Carroll, S. R., & Martinez, A. (2021). Indigenous data sovereignty and policy. Abingdon: Taylor & Francis.