

Transparência no Relato de Aspectos Éticos e de Proteção de Dados em Computação na Saúde com Dados Clínicos e Não Clínicos: Análise do SBCAS 2025

Maria da Graça C. Pimentel¹, Érika H.P. Cardoso¹, Sincler P. Meireles¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo

mgp@icmc.usp.br, erika.cardoso@alumni.usp.br, sincler@usp.br

Abstract. *Digital and social data potentially associated with individuals constitute personal data subject to legal and ethical protection. In this context, this paper presents an exploratory analysis of the explicit reporting of ethical and data protection aspects in Health Computing research, based on papers published at SBCAS 2025 involving clinical and non-clinical data. A manual review supported by a computational pipeline was conducted to identify references to ethical governance, consent, and data protection. The results indicate variability in how ethical and data protection practices are reported, especially in studies involving digital and non-clinical data, as well as gaps in transparency regarding ethical governance and personal data handling.*

Resumo. *Dados digitais e sociais potencialmente associados a indivíduos constituem dados pessoais sujeitos a proteção legal e ética. Nesse contexto, este artigo apresenta uma análise exploratória da explicitação de aspectos éticos e de proteção de dados em pesquisas em Computação aplicada à saúde, com base em artigos do SBCAS 2025 envolvendo dados clínicos e não clínicos. Utiliza-se revisão manual apoiada por pipeline computacional para identificar referências a governança ética, consentimento e proteção de dados. Os resultados indicam variabilidade na forma de reportar práticas éticas e de proteção de dados, especialmente em estudos com dados digitais e não clínicos, além de lacunas na transparência sobre governança ética e tratamento de dados pessoais.*

1. Introdução

O crescente uso de dados digitais e sociais tem ampliado significativamente as possibilidades de pesquisa em saúde, especialmente no contexto da Computação. Plataformas online, redes sociais e outros ecossistemas digitais passaram a ser utilizados como fontes de dados comportamentais, permitindo a análise de percepções e riscos associados à saúde individual e coletiva. Esse movimento se manifesta particularmente em estudos baseados em dados digitais não clínicos, o que tem sido acompanhado pela consolidação de uma agenda de pesquisa específica e pelo surgimento de iniciativas dedicadas ao tema, como o *workshop Mining Digital and Social Signals for Public Health* (MinDS) organizado por Ferreira and Lima Filho [2026] no contexto do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS) 2026. Paradoxalmente, embora esses estudos frequentemente utilizem dados pessoais potencialmente associados a indivíduos — ainda que publicamente acessíveis — aspectos relacionados à governança ética, consentimento e proteção de dados permanecem pouco explicitados. Nesse sentido, o contexto do MinDS também

evidencia a relevância de discutir transparência ética e proteção de dados em pesquisas baseadas em sinais digitais e sociais aplicados à saúde.

Ao mesmo tempo, a literatura já vem, há mais de uma década, chamando atenção para desafios associados ao uso desses dados digitais,¹ incluindo questões relacionadas à qualidade, representatividade e implicações éticas de sua utilização. Boyd and Crawford [2012] destacam a necessidade de uma análise crítica dos pressupostos e vieses associados ao uso de Big Data. De forma complementar, Tufekci [2014] discute desafios metodológicos e conceituais desse campo emergente, com ênfase na validade e representatividade de análises baseadas em dados de mídias sociais. Por sua vez, Zook et al. [2017] evidenciam riscos éticos concretos associados ao uso desses dados, incluindo pesquisas altamente invasivas baseadas no cruzamento de dados públicos, citando exemplos como o experimento de contágio emocional realizado pelo Facebook e o uso de técnicas de análise geográfica para inferir a identidade do artista conhecido pelo pseudônimo Banksy.

Nesse contexto, dados digitais não são eticamente neutros — mesmo quando publicamente acessíveis. Metcalf and Crawford [2016] problematizam a própria noção de “sujeitos de pesquisa” em estudos baseados em Big Data, argumentando que indivíduos frequentemente são tratados como meras fontes de dados, sem reconhecimento explícito de sua condição de participantes. Já Fiesler and Proferes [2018] evidenciam, a partir da perspectiva dos próprios usuários, que há um desalinhamento entre as expectativas dos participantes e o uso de seus dados em pesquisas, incluindo preocupações com privacidade, consentimento e reidentificação.

Em nível internacional, evidências empíricas recentes indicam que a explicitação de aspectos éticos e de proteção de dados em pesquisas baseadas em mídias sociais permanece limitada. Como exemplo, a *scoping review* de Zhang et al. [2024], envolvendo 108 estudos sobre mineração de textos em comunidades online, observou que apenas 59,3% reportavam aprovação ética e somente 12,9% explicitavam a obtenção efetiva de consentimento informado. Os autores também destacam o uso frequente de conteúdo textual de usuários, muitas vezes com citações diretas potencialmente rastreáveis.

No contexto brasileiro, a proteção de dados pessoais é respaldada por um conjunto de instrumentos normativos que incluem a Constituição Federal de 1988 [Brasil, 1988], a Lei Geral de Proteção de Dados Pessoais (LGPD) [Brasil, 2018]² e diretrizes específicas para pesquisa envolvendo seres humanos, como a Resolução nº 510 do Conselho Nacional de Saúde [Brasil, 2016] e orientações da Autoridade Nacional de Proteção de Dados (ANPD) [ANPD, 2022]. Esse arcabouço normativo estabelece princípios como finalidade, necessidade, transparência e responsabilização no tratamento de dados, incluindo sua aplicação em contextos de pesquisa acadêmica.

Apesar dessa estrutura normativa, a operacionalização desses princípios em pesquisas envolvendo dados digitais apresenta desafios. Carvalho et al. [2019] discutem dificuldades relacionadas à transparência no tratamento de dados em cenários marcados

¹Neste trabalho, dados digitais não clínicos incluem, entre outros, dados provenientes de redes sociais e plataformas online.

²LGPD Art. 5, I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável; II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural.

pelo uso intensivo de algoritmos e grandes volumes de informação. Carvalho et al. [2020] analisam especificamente pesquisas em análise de redes sociais, destacando a necessidade de abordagens interdisciplinares e boas práticas para o tratamento de dados pessoais conforme a LGPD. A discussão de Carvalho et al. [2021] incorpora aspectos éticos, jurídicos e computacionais relacionados ao uso de dados pessoais em pesquisa.

Diante desse cenário, observa-se uma lacuna específica relacionada à transparência no relato de aspectos éticos e de proteção de dados em pesquisas em saúde. Embora estudos internacionais evidenciem limitações na explicitação dessas dimensões em pesquisas baseadas em dados digitais nessa área, no Brasil ainda há pouca articulação entre aspectos éticos e proteção de dados, o que é particularmente crítico no contexto da saúde. Assim, torna-se relevante analisar essas dimensões de forma integrada nesse domínio.

Nesse contexto, este trabalho busca contribuir para a discussão dessas dimensões por meio de uma análise dos artigos publicados no SBCAS 2025, escolhido por sua relevância como principal fórum nacional na área de Computação Aplicada à Saúde e por representar um recorte recente das práticas de explicitação de aspectos éticos, governança de dados e proteção de dados pessoais na área. Essa escolha metodológica restringe a análise a uma única edição do evento e deixa de contemplar trabalhos de anos anteriores que explicitam aprovação ética, como o estudo de Montero and Conci [2024].

Utiliza-se uma revisão manual, apoiada por um pipeline computacional de detecção de sinais textuais, com o objetivo de identificar a presença de referências a governança ética, consentimento e proteção de dados, incluindo anonimização. Diferentemente de abordagens focadas na verificação de conformidade, este estudo concentra-se na análise da explicitação desses aspectos nos textos, permitindo avaliar padrões de transparência no relato dessas práticas em pesquisas em saúde, em particular naquelas que utilizam dados digitais não clínicos.

No restante deste texto, a Seção 2 discute trabalhos relacionados, a Seção 3 sintetiza a metodologia adotada, a Seção 4 sumariza e discute os resultados obtidos, a Seção 5 contribui com um conjunto de recomendações para artigos da área que reportam o uso de dados pessoais, e a Seção 6 apresenta considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Estudos empíricos têm investigado a explicitação de aspectos éticos e de proteção de dados em pesquisas que utilizam dados digitais em saúde há mais de uma década. Por exemplo, Golder et al. [2017] analisaram estudos baseados em mídias sociais, incluindo 17 estudos qualitativos, e identificaram limitações no reporte de aprovação ética, consentimento informado e medidas de proteção à privacidade.

De forma complementar, Takats et al. [2022] revisaram 367 estudos que utilizam dados do Twitter em saúde pública e observaram que apenas cerca de um terço reportava aprovação ética, além da presença recorrente de informações potencialmente identificáveis e de estratégias limitadas de anonimização.

Mais recentemente, Zhang et al. [2024] conduziram uma *scoping review* com 108 estudos sobre mineração de textos em comunidades online, abrangendo 19 países de cinco continentes. Os autores observaram que apenas 59,3% dos estudos reportavam aprovação ética, 45,3% mencionavam consentimento informado e somente 12,9% explicitavam sua

obtenção efetiva. Aproximadamente 70% dos estudos continham conteúdo textual produzido por usuários, sendo frequente o uso de citações diretas potencialmente rastreáveis. Embora parte dos trabalhos recorresse à paráfrase ou agregação dos dados como estratégia de mitigação de riscos, nem sempre os procedimentos adotados eram explicitados.

Resultados recentes reforçam esse cenário. Sadat et al. [2025] analisaram 82 estudos que utilizam netnografia em saúde e observaram variação significativa na transparência ética: apenas 33 estudos reportaram aprovação ética, enquanto pouco mais da metade mencionou consentimento informado. No contexto da análise computacional de dados em saúde pública, Villanueva-Miranda et al. [2025] revisaram 83 estudos sobre análise de sentimentos, destacando que o uso predominante de dados de mídias sociais traz desafios relacionados à qualidade dos dados, vies e questões éticas.

Em conjunto, esses resultados evidenciam limitações recorrentes na explicitação de aprovação ética, consentimento informado e estratégias de proteção de dados pessoais em pesquisas baseadas em dados digitais em saúde.

3. Método

Para o levantamento, foram analisados os 96 artigos publicados no SBCAS no ano de 2025 (82 artigos completos e 14 artigos curtos) obtidos da SBC Open Library (SOL).³

A partir desse conjunto, buscou-se construir uma visão geral da ocorrência e explicitação de aspectos éticos e de proteção de dados nas publicações dessa edição do evento. A análise combinou busca textual apoiada por um pipeline computacional de detecção de sinais textuais com revisão manual. O pipeline auxiliou a identificação inicial de elementos como referências a CEP, TCLE e participação humana, enquanto a classificação final das dimensões analisadas foi realizada manualmente. O conjunto inicial de sinais textuais foi refinado iterativamente durante a análise manual (Apêndice A).

As dimensões analisadas neste estudo foram definidas com base em elementos centrais das normativas brasileiras aplicáveis à pesquisa envolvendo dados e seres humanos, em especial a LGPD e a Resolução CNS nº 510/2016. Em particular, foram consideradas como evidências de explicitação textual aspectos como: (i) governança ética, incluindo submissão ou dispensa por Comitê de Ética em Pesquisa (CEP); (ii) consentimento, incluindo a obtenção de Termo de Consentimento Livre e Esclarecido (TCLE), quando aplicável; e (iii) proteção de dados, incluindo menções a anonimização, tratamento de dados pessoais e conformidade com a LGPD.

Dessa forma, as strings utilizadas nas buscas, tanto pelo pipeline computacional como manualmente, operacionalizam, no nível textual, a presença ou ausência de sinais textuais associados a essas dimensões. O objetivo não é avaliar conformidade legal, mas identificar o grau de explicitação desses aspectos nos artigos analisados.

A etapa de análise envolveu a aplicação sistemática dessas strings em todos os artigos do corpus — em uma primeira etapa com o apoio do pipeline. A presença de cada dimensão foi considerada quando ao menos uma das expressões associadas foi identificada de forma explícita no texto. Cada ocorrência foi analisada considerando o contexto correspondente, buscando identificar menções explícitas e objetivas aos aspectos investigados.

³<https://sol.sbc.org.br/index.php/sbcas/issue/view/1514>

A partir dessa análise, os artigos foram classificados conforme a presença ou ausência desses elementos, resultando em uma visão quantitativa do estado das publicações em relação às dimensões éticas consideradas.

4. Resultados e Discussão

A Tabela 1 caracteriza os estudos segundo origem e disponibilização dos dados.

Código	Categoria	n
SLR	Revisões sistemáticas da literatura	6
TSS	Estudos teóricos/simulações/dados sintéticos sem dados pessoais	5
TP-GOV-only	Exclusivamente dataset governamental	5
TP-GOV+	Dataset governamental + outras fontes	9
TP-ND	Dataset de terceiros não disponibilizado publicamente	10
TP-D	Dataset de terceiros disponibilizado publicamente	50
AU-PREV-ND	Dataset prévio dos autores não disponibilizado publicamente	3
AU-PREV-D	Dataset prévio dos autores disponibilizado publicamente	3
AU-NEW	Dataset inédito produzido pelos autores para o estudo	37

Tabela 1. Caracterização dos estudos segundo origem e disponibilização dos dados: categorias não mutuamente exclusivas.

Entre os 96 artigos analisados, seis correspondem a revisões sistemáticas da literatura (SLR, n=6), as quais, conforme a Resolução CNS nº 510/2016, podem ser dispensadas de avaliação por CEP por não envolverem interação direta com participantes nem coleta de dados identificáveis. Além disso, cinco artigos não envolvem dados pessoais por realizarem investigações teóricas, simulações ou utilizarem dados sintéticos (TSS, n=5).

Um total de 14 estudos fizeram uso, exclusivamente ou não, de *datasets* disponibilizados por órgãos governamentais, os quais, em princípio, são disponibilizados sob critérios de anonimização e agregação. Embora esses dados sejam frequentemente considerados públicos, seu uso em pesquisa não elimina a necessidade de considerar limitações de finalidade e possíveis riscos de reidentificação.

Observa-se predominância de estudos utilizando *datasets* previamente disponibilizados publicamente por terceiros (TP-D, n=50), bem como presença relevante de dados produzidos pelos próprios autores, incluindo *datasets* inéditos produzidos especificamente para os estudos analisados (AU-NEW, n=37) e *datasets* previamente produzidos pelos autores, disponibilizados publicamente (AU-PREV-D, n=3) ou não disponibilizados publicamente (AU-PREV-ND, n=3).

Em geral, nos estudos baseados em *datasets* previamente disponibilizados ou reutilizados, não se observa discussão explícita sobre contexto original de coleta, limitações de reutilização ou possíveis implicações éticas e de proteção de dados associadas ao uso secundário dessas informações, mesmo quando publicamente acessíveis. Entre os trabalhos que produziram *datasets* inéditos, apenas 11 apresentaram alguma discussão relacionada ao tratamento de dados pessoais, incluindo aspectos como coleta, armazenamento, anonimização ou compartilhamento, enquanto 25 não apresentaram qualquer menção a esse tipo de cuidado.

A Tabela 2 sumariza características potencialmente sensíveis relacionadas ao tratamento de dados identificadas no corpus analisado.

Código	Característica	n
SOCIAL	Uso de dados de redes sociais	3
IMG	Coleta de imagens	7
MEX	Uso de dispositivos de monitoramento externo	5
BIO-COL	Coleta de amostras biológicas	2
BIO-BNC	Uso de material biológico armazenado	1
NONINV	Coleta dirigida de dados não invasiva	17
NO-DISC	Sem discussão sobre tratamento de dados pessoais	66

Tabela 2. Características potencialmente sensíveis relacionadas ao tratamento de dados.

Foram identificados estudos envolvendo categorias potencialmente sensíveis de dados, incluindo coleta de imagens (IMG, n=7), uso de dispositivos de monitoramento externo (MEX, n=5), coleta de amostras biológicas (BIO-COL, n=2) e uso de dados de redes sociais (SOCIAL, n=3). No total, 66 artigos não apresentaram discussão explícita sobre tratamento ou proteção de dados pessoais.

No caso específico dos estudos envolvendo redes sociais, não se observa explicitação de aspectos relacionados a consentimento, riscos de reidentificação ou implicações éticas associadas ao uso desses dados, apesar de envolverem dados digitais potencialmente associados a indivíduos.

Um total de 43 estudos fizeram uso de dados coletados pelos próprios autores (AU-NEW, n=37; AU-PREV-D, n=3; AU-PREV-ND, n=3). Entretanto, apenas 14 artigos registram explicitamente aprovação prévia por CEP para o estudo reportado, enquanto apenas dois apresentam justificativa explícita de dispensa de apreciação ética. Esse resultado evidencia uma lacuna significativa na transparência quanto aos procedimentos de governança ética adotados nas pesquisas analisadas.

Observa-se ainda que apenas dois dos 96 artigos analisados apresentam seção específica dedicada à discussão de aspectos éticos. Esse cenário tende a se modificar, considerando que eventos da área vêm adotando exigências explícitas nesse sentido, como o IHC⁴ – Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais/SBC, que passou a exigir a seção de Cuidados Éticos a partir de 2025 e descrição de medidas de proteção de dados pessoais em 2026, e o WebMedia⁵ – Simpósio Brasileiro de Sistemas Multimídia e Web, que passou a exigir, a partir de 2026, seção específica sobre Cuidados Éticos e Atendimento à LGPD.

Considerando o perfil dos estudos publicados no SBCAS e em seus eventos satélites, frequentemente envolvendo dados clínicos, comportamentais, biométricos ou potencialmente associados a indivíduos, seria importante a adoção de exigências semelhantes também nesse contexto, favorecendo maior transparência quanto à governança ética e ao tratamento de dados pessoais nas pesquisas publicadas.

De forma geral, os resultados evidenciam um cenário no qual o uso de dados clínicos e não clínicos é amplamente disseminado, enquanto a explicitação de aspectos éticos e de proteção de dados permanece limitada, particularmente em estudos envolvendo reutilização de dados digitais e uso secundário de informações publicamente acessíveis.

⁴<https://ihc.sbc.org.br/2026/index.php/artigos-completos/>

⁵https://webmedia.org.br/2026/?page_id=38

5. Recomendações

Considerando a legislação brasileira vigente e o Código⁶ de Conduta para Autores em Publicações da SBC, em especial o Art. 2º IV (pesquisas envolvendo seres humanos) e V (pesquisas envolvendo dados pessoais), e com base nos resultados observados, propõem-se as recomendações apresentadas na Tabela 3 para artigos da área de Computação Aplicada à Saúde — como os publicados no SBCAS e em eventos correlatos — que reportam pesquisas envolvendo o uso de dados, especialmente dados digitais e não clínicos:

ID	Recomendação
R1	Inclusão de seção específica sobre aspectos éticos e de proteção de dados. Recomenda-se que artigos incluam uma seção dedicada à descrição de cuidados éticos, contemplando de forma explícita a governança da pesquisa, o tratamento de dados e eventuais riscos aos participantes.
R2	Transparência quanto à avaliação por CEP. Estudos que envolvam dados potencialmente associados a indivíduos devem explicitar claramente se foram submetidos a Comitê de Ética em Pesquisa (CEP), ou justificar sua dispensa conforme a Resolução CNS nº 510/2016.
R3	Especificação da base legal, finalidade e aspectos críticos do tratamento de dados. Pesquisas que utilizam dados pessoais devem indicar de forma explícita a finalidade do tratamento e a base legal correspondente, em conformidade com a LGPD, mesmo quando os dados forem considerados públicos. Adicionalmente, recomenda-se explicitar, de forma concisa, aspectos do tratamento de dados que sejam relevantes do ponto de vista ético e de proteção de dados, como: (i) forma de obtenção dos dados; (ii) existência de etapas de inferência ou uso em modelos que possam impactar a privacidade dos indivíduos; (iii) eventual compartilhamento ou divulgação de dados ou resultados; e (iv) critérios de retenção e descarte. Não se trata de detalhar toda a metodologia, mas de tornar explícitos os pontos críticos para avaliação ética e legal do estudo.
R4	Descrição das medidas de proteção de dados adotadas. Recomenda-se explicitar as medidas concretas adotadas para proteção dos dados ao longo do estudo, incluindo, quando aplicável, procedimentos de anonimização, pseudonimização, agregação ou controle de acesso. Essa descrição deve permitir compreender como os riscos identificados foram efetivamente mitigados.
R5	Consideração ética no uso de <i>datasets</i> de terceiros. O uso de dados disponibilizados publicamente não elimina a necessidade de avaliação ética. Deve-se discutir as condições de coleta original dos dados e os limites de sua reutilização.
R6	Explicitação de práticas de consentimento. Quando aplicável, a obtenção de consentimento deve ser claramente descrita. Em casos em que não há interação direta com participantes, recomenda-se justificar a não aplicabilidade.
R7	Definição de políticas de armazenamento, retenção e descarte de dados. Recomenda-se explicitar como os dados são armazenados ao longo da pesquisa, incluindo aspectos de segurança e controle de acesso, bem como definir critérios para retenção e descarte após o término do estudo. Essa definição deve considerar a minimização de dados e a limitação de uso ao período necessário para os objetivos da pesquisa.
R8	Integração entre ética e proteção de dados. Recomenda-se que governança ética, consentimento e proteção de dados não sejam tratados de forma isolada, mas como dimensões interdependentes do mesmo processo de pesquisa.

Tabela 3. Recomendações para explicitação de aspectos éticos e proteção de dados em pesquisas na área.

⁶<https://sol.sbc.org.br/index.php/indice/conduta>

Essas recomendações visam promover maior transparência, reprodutibilidade e responsabilidade no uso de dados em pesquisas em saúde, alinhando práticas da área aos requisitos normativos vigentes e às expectativas éticas contemporâneas.

6. Conclusão

Este trabalho apresentou uma análise da explicitação de aspectos éticos e de proteção de dados em artigos publicados no SBCAS 2025. A partir da análise de 96 artigos, observou-se um uso amplo e diversificado de dados, incluindo *datasets* de terceiros, dados governamentais, dados produzidos pelos próprios autores e dados extraídos de redes sociais.

Os resultados evidenciam uma baixa explicitação de práticas relacionadas à governança ética, consentimento e proteção de dados. Em particular, verificou-se a ausência de seções específicas dedicadas a aspectos éticos e de proteção de dados na maioria dos artigos, bem como a falta de menção à avaliação por Comitê de Ética em Pesquisa (CEP), à obtenção de consentimento e à conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD). Esses achados indicam uma lacuna relevante entre o uso efetivo de dados em pesquisas em saúde e a explicitação das práticas éticas e legais associadas a esse uso. Tal lacuna sugere a necessidade de maior alinhamento entre as práticas da área e os requisitos normativos vigentes: o conjunto de recomendações apresentado visa fomentar essa explicitação.

Este estudo apresenta algumas limitações, incluindo a análise restrita à edição mais recente do SBCAS. Além disso, embora o pipeline computacional já permita identificar sinais relacionados a aspectos éticos e proteção de dados, sua capacidade analítica ainda é limitada por abordagens predominantemente baseadas em regras e padrões textuais. Como continuidade, o pipeline está sendo estendido com mecanismos adicionais de análise textual, buscando ampliar a capacidade de identificação e categorização dos elementos analisados sem comprometer auditabilidade, rastreabilidade e interpretabilidade das análises produzidas.

Como trabalhos futuros, pretende-se ampliar a análise para outras edições do SBCAS e eventos correlatos, bem como investigar de forma mais aprofundada a relação entre tipos de dados utilizados e o nível de explicitação de aspectos éticos e de proteção de dados, contribuindo para o desenvolvimento de diretrizes estruturadas para a área.

7. Agradecimentos

Agradecemos às pessoas pesquisadoras Luiz Paulo Carvalho, Kamila Rios da Hora Rodrigues e André Pimenta Freire por seus trabalhos e parcerias inspiradores.

8. Aspectos éticos e de proteção de dados

Os autores utilizaram o sistema ChatGPT (OpenAI) exclusivamente como apoio à revisão textual. Não houve coleta ou tratamento de dados pessoais. Considerando que o corpus consiste em artigos científicos de acesso público, sem interação com participantes, a pesquisa não se enquadra como pesquisa envolvendo seres humanos, conforme a Resolução CNS nº 510/2016. Ainda assim, foram observados os princípios da LGPD, em especial finalidade, minimização e transparência.

Referências

- ANPD. Guia Orientativo: Tratamento de Dados Pessoais para Fins Acadêmicos e de Pesquisa. <https://www.gov.br/anpd/pt-br/centrais-de-conteudo/materiais-educativos-e-publicacoes/web-guia-anpd-tratamento-de-dados-para-fins-academicos.pdf>, 2022.
- Danah Boyd and Kate Crawford. Critical Questions for Big Data. *Information, Communication & Society*, 15(5):662–679, 2012. doi: 10.1080/1369118X.2012.678878.
- Brasil. Constituição da República Federativa do Brasil de 1988. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm, 1988.
- Brasil. Resolução nº 510, de 7 de abril de 2016. Conselho Nacional de Saúde. <https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2016/resolucao-no-510.pdf>, 2016.
- Brasil. Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais – LGPD). https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709compilado.htm, 2018.
- Luiz Paulo Carvalho, J. Oliveira, Claudia Cappelli, and V. Majer. Desafios de Transparência pela Lei Geral de Proteção de Dados Pessoais. In *Anais do SBSI*, 2019.
- Luiz Paulo Carvalho, Jonice Oliveira, and Claudia Cappelli. Pesquisas em Análise de Redes Sociais e LGPD, análises e recomendações. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 73–84, Porto Alegre, RS, Brasil, 2020. SBC. doi: 10.5753/brasnam.2020.11164.
- Luiz Paulo Carvalho, Jonice Oliveira, Flávia Maria Santoro, and Claudia Cappelli. Social Network Analysis, Ethics and LGPD, considerations in research. *iSys - Journal of Information Systems*, 14(2):28–52, Aug. 2021. doi: 10.5753/isys.2021.1235.
- Carlos Henrique Gomes Ferreira and Silas Pereira Lima Filho. I Workshop on Mining Digital and Social Signals for Public Health (MinDS). <https://www.sbcas2026.com/workshop-mind-ds>, 2026. Workshop at SBCAS 2026.
- Casey Fiesler and Nicholas Proferes. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society*, 4(1), 2018. doi: 10.1177/2056305118763366.
- Su Golder, Saad Ahmed, Gillian Norman, and Andrew Booth. Attitudes toward the ethics of research using social media: A systematic review. *Journal of Medical Internet Research*, 19(6):e195, 2017. doi: 10.2196/jmir.7082.
- Jacob Metcalf and Kate Crawford. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. *Big Data & Society*, 3(1), 2016. doi: 10.1177/2053951716650211.
- José Montero and Aura Conci. Estudo relacionado à identificação da malignidade de nódulos da tireóide por dados clínicos e imagens infravermelhas. In *Anais Estendidos do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 103–108, Porto Alegre, RS, Brasil, 2024. SBC. doi: 10.5753/sbcas_estendido.2024.2247.
- Amany Sadat, Elizabeth Green, Imogen Forsythe, Stacey Munnelly, Georgette Eaton, Matthew Wynn, Fiona Pearson, and Emma Dobson. Utilization of Netnography as a Health Care Research Methodology: Scoping Review. *Journal of Medical Internet Research*, 27:e78025, 2025. doi: 10.2196/78025.

C Takats, A Kwan, R Wormer, D Goldman, H Jones, and D Romero. Ethical and Methodological Considerations of Twitter Data for Public Health Research: Systematic Review. *Journal of Medical Internet Research*, 24(11):e40380, 2022. doi: 10.2196/40380.

Zeynep Tufekci. Big Data and Social Science: Three Objections. *Sociological Inquiry*, 84(4):505–512, 2014. doi: 10.1111/soin.12065.

Ismael Villanueva-Miranda, Yang Xie, and Guanghua Xiao. Sentiment analysis in public health: a systematic review of the current state, challenges, and future directions. *Frontiers in Public Health*, Volume 13 - 2025, 2025. ISSN 2296-2565. doi: 10.3389/fpubh.2025.1609749.

Y. Zhang, J. Fu, J. Lai, S. Deng, Z. Guo, C. Zhong, J. Tang, W. Cao, and Y. Wu. Reporting of Ethical Considerations in Qualitative Research Utilizing Social Media Data on Public Health Care: Scoping Review. *Journal of Medical Internet Research*, 26:e51496, 2024. doi: 10.2196/51496.

Matthew Zook, Solon Barocas, Danah Boyd, Kate Crawford, Emily Keller, Seeta P. Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, and Alondra Nelson. Ten Simple Rules for Responsible Big Data Research. *PLoS Computational Biology*, 13(3):e1005399, 2017. doi: 10.1371/journal.pcbi.1005399.

A. Dicionário de Sinais para Detecção de Aspectos Éticos e de Proteção de Dados

O dicionário inicial (Tabela 4) de sinais, estendido durante a análise a partir de ocorrências observadas no corpus, foi estruturado em quatro categorias principais, cada uma composta por termos amplos e expressões estritas.

Categoria	Termos amplos	Expressões estritas
Governança ética	ética; ethics; comitê; committee; IRB; CEP; CONEP; CAAE; SI-NEP; ANPD	comitê de ética em pesquisa; Plataforma Brasil; approved by the ethics committee; institutional review board approval; Resolução CNS 466; Resolução CNS 510; sistema CEP/CONEP; relatório de impacto à proteção de dados pessoais
Consentimento	consentimento; consent; TCLE; TALE; assentimento; voluntário	termo de consentimento livre e esclarecido; informed consent was obtained; participação voluntária; consent form; consent from the human subjects
Tratamento e anonimização	anonimiz*; anonymiz*; pseudonimiz*; LGPD; GDPR; privacidade; confidentiality; desidentificação	dados foram anonimizados; data was anonymized; personally identifiable information was removed; lei geral de proteção de dados; dados pseudonimizados
Natureza dos dados	dataset; conjunto de dados; dados pessoais; dados sensíveis; participantes; seres humanos; human subjects	dados acadêmicos; dados pessoais sensíveis; pesquisa envolvendo seres humanos; personally identifiable information; dados de saúde

Tabela 4. Categorias e exemplos de termos utilizados na identificação de aspectos éticos e de proteção de dados.