

QUEM O ALGORITMO ENXERGA? DIVERSIDADE DEMOGRÁFICA E VALIDAÇÃO EXTERNA DE MODELOS DE INTELIGÊNCIA ARTIFICIAL APLICADOS À ANÁLISE ESTÉTICA FACIAL E DO SORRISO: UMA REVISÃO SISTEMÁTICA METODOLÓGICA

WHO DOES THE ALGORITHM SEE? DEMOGRAPHIC DIVERSITY AND EXTERNAL VALIDATION OF ARTIFICIAL INTELLIGENCE MODELS APPLIED TO FACIAL AND SMILE ESTHETIC ANALYSIS: A METHODOLOGICAL SYSTEMATIC REVIEW

¿A QUIÉN VE EL ALGORITMO? DIVERSIDAD DEMOGRÁFICA Y VALIDACIÓN EXTERNA DE MODELOS DE INTELIGENCIA ARTIFICIAL APLICADOS AL ANÁLISIS ESTÉTICO FACIAL Y DE LA SONRISA: UNA REVISIÓN SISTEMÁTICA METODOLÓGICA

Daniel da Penha Soares¹

Rim Hassan Youssef²

Ana Valéria Pagliari³

RESUMO: Este artigo buscou avaliar o relato da diversidade demográfica e a realização de validação externa em estudos sobre modelos de inteligência artificial (IA) aplicados à análise estética facial e do sorriso. Trata-se de uma revisão sistemática metodológica, com buscas nas bases PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO em outubro de 2026, relatada conforme o PRISMA 2020 e restrita a estudos com texto completo disponível gratuitamente. Dos 264 registros únicos, 14 estudos foram incluídos, publicados entre 2022 e 2026 e provenientes de 12 países. O sexo dos participantes foi relatado em 78,6% dos estudos, a idade em 57,1% e a etnia ou ancestralidade em 35,7%; entre estes, três de cinco utilizaram amostras de uma única população. Nenhum dos oito modelos desenvolvidos pelos autores passou por validação externa, e apenas uma das seis ferramentas prontas e modelos de linguagem teve a população de treinamento descrita. Nenhum estudo avaliou pacientes brasileiros. Conclui-se que os modelos de IA para análise estética são desenvolvidos e testados em populações pouco descritas e pouco diversas, o que limita sua generalização para países miscigenados, como o Brasil, e reforça a necessidade de relato padronizado e de validação externa antes do uso clínico.

Palavras-chave: Inteligência artificial. Estética dentária. Ortodontia. Grupos raciais. Viés.

¹Graduando em Odontologia (discente do curso de graduação em Odontologia), Centro Universitário Dinâmica das Cataratas (UDC), Foz do Iguaçu, PR, Brasil.

²Graduanda em Odontologia (discente do curso de graduação em Odontologia), Centro Universitário Dinâmica das Cataratas (UDC), Foz do Iguaçu, PR, Brasil.

³Orientadora. Doutora em Odontologia Preventiva e Social; docente do curso de graduação em Odontologia e Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Araçatuba, SP, Brasil. Função acadêmica: Centro Universitário Dinâmica das Cataratas (UDC), Foz do Iguaçu, PR, Brasil.

ABSTRACT: This article aimed to assess the reporting of demographic diversity and the performance of external validation in studies on artificial intelligence (AI) models applied to facial and smile esthetic analysis. This is a methodological systematic review, with searches of PubMed/MEDLINE, Europe PMC, OpenAlex and SciELO in October 2026, reported according to PRISMA 2020 and restricted to studies with freely available full texts. Of the 264 unique records, 14 studies were included, published between 2022 and 2026 and originating from 12 countries. Participants' sex was reported in 78.6% of the studies, age in 57.1% and ethnicity or ancestry in 35.7%; among these, three of five used samples from a single population. None of the eight models developed by the authors underwent external validation, and only one of the six ready-made tools and language models had its training population described. No study assessed Brazilian patients. AI models for esthetic analysis are developed and tested in poorly described and poorly diverse populations, which limits their generalizability to admixed countries such as Brazil and reinforces the need for standardized reporting and external validation before clinical use.

Keywords: Artificial intelligence. Esthetics. Dental. Orthodontics. Racial groups. Bias.

RESUMEN: Este artículo evaluó el reporte de la diversidad demográfica y la realización de validación externa en estudios sobre modelos de inteligencia artificial (IA) aplicados al análisis estético facial y de la sonrisa. Revisión sistemática metodológica con búsquedas en PubMed/MEDLINE, Europe PMC, OpenAlex y SciELO en octubre de 2026, reportada según PRISMA 2020 y restringida a estudios con texto completo gratuito. De los 264 registros únicos, se incluyeron 14 estudios, publicados entre 2022 y 2026 y procedentes de 12 países. El sexo de los participantes se informó en el 78,6% de los estudios, la edad en el 57,1% y la etnia o ascendencia en el 35,7%; entre estos, tres de cinco utilizaron muestras de una única población. Ninguno de los ocho modelos desarrollados por los autores fue sometido a validación externa, y solo una de las seis herramientas listas y modelos de lenguaje tuvo descrita su población de entrenamiento. Ningún estudio evaluó pacientes brasileños. Se concluye que los modelos de IA para análisis estético se desarrollan y prueban en poblaciones poco descritas y poco diversas, lo que limita su generalización a países mestizos, como Brasil, y refuerza la necesidad de un reporte estandarizado y de validación externa antes del uso clínico.

Palabras clave: Inteligencia artificial. Estética dental. Ortodoncia. Grupos raciales. Sesgo.

INTRODUÇÃO

A inteligência artificial (IA) tem sido aplicada à análise estética da face e do sorriso para atribuir escores de atratividade, estimar a idade aparente, classificar o perfil facial e avaliar índices estéticos gengivais e dentários (OBWEGESER D, et al., 2022; MA Q, et al., 2024; ELVERIŞLI Ö, et al., 2026; IBRAHIM M, et al., 2026). A promessa dessas ferramentas é trazer objetividade e reprodutibilidade a um campo tradicionalmente subjetivo, apoiando o diagnóstico, a comunicação com o paciente e a avaliação de resultados ortodônticos e cirúrgicos.

Entretanto, os modelos de IA aprendem a partir dos dados usados em seu desenvolvimento, e seu desempenho depende de quão bem esses dados representam a população em que serão aplicados. Na medicina, um algoritmo amplamente utilizado nos Estados Unidos subestimou as necessidades de saúde de pacientes negros porque aprendeu padrões enviesados presentes nos dados de treinamento (OBERMEYER Z, et al., 2019). Na

estética, o risco é particularmente relevante, pois a percepção de beleza e as proporções faciais variam entre grupos étnicos, sexos e idades, e um modelo treinado com rostos de uma única população tende a reproduzir o padrão estético dessa população.

Uma revisão recente sobre IA na ortodontia apontou três problemas recorrentes em todas as áreas analisadas: dados de treinamento restritos a um único centro, representação demográfica estreita e ausência de validação externa independente. A mesma revisão relatou que modelos de decisão sobre extrações perderam até 20% de sua acurácia quando testados em outras instituições (ALKADHI OH, 2026). Diretrizes de relato de modelos preditivos baseados em aprendizado de máquina recomendam descrever as fontes de dados e as características dos participantes, informar como o desempenho foi avaliado e considerar a equidade do modelo entre subgrupos (COLLINS GS, et al., 2024).

No Brasil, essa discussão ganha importância adicional. Segundo o Censo Demográfico de 2022, 45,3% da população se declarou parda, 43,5% branca, 10,2% preta, 0,8% indígena e 0,4% amarela (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2023). Ferramentas desenvolvidas com dados de outras populações podem não representar adequadamente essa diversidade, mas não se sabe em que medida os estudos sobre IA aplicada à estética facial e do sorriso descrevem as populações utilizadas e testam seus modelos em dados independentes.

Diante dessa lacuna, este artigo buscou avaliar o relato da diversidade demográfica (sexo, idade e etnia ou ancestralidade) e a realização de validação externa nos estudos sobre modelos de IA aplicados à análise estética facial e do sorriso.

MÉTODOS

DESENHO DO ESTUDO

Trata-se de uma revisão sistemática metodológica, também denominada meta-pesquisa, cujo objeto é a forma como os estudos foram conduzidos e relatados, e não a eficácia das intervenções avaliadas. O estudo seguiu as recomendações para pesquisas metaepidemiológicas (MURAD MH e WANG Z, 2017), e o processo de seleção foi relatado conforme o PRISMA 2020 (PAGE MJ, et al., 2021). A pergunta norteadora foi: como os estudos sobre modelos de IA aplicados à análise estética facial e do sorriso relatam a diversidade demográfica de suas amostras e se realizam validação externa?

FONTES DE DADOS E ESTRATÉGIA DE BUSCA

As buscas foram realizadas em outubro de 2026 nas bases PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO, com filtro de publicação de 2016 a 2026. A estratégia combinou três blocos de termos, referentes à IA, à estética facial e do sorriso e ao contexto odontológico, adaptados à sintaxe de cada base. Para ampliar a cobertura da literatura ibero-americana, a busca na SciELO combinou termos em português, inglês e espanhol referentes à IA e à estética, e foi realizada uma busca complementar no OpenAlex com termos equivalentes em português e espanhol. A estratégia utilizada no PubMed/MEDLINE foi a seguinte:

("deep learning"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "neural network"[Title/Abstract] OR "neural networks"[Title/Abstract] OR "artificial intelligence"[Title/Abstract] OR "convolutional"[Title/Abstract]) AND ("facial attractiveness"[Title/Abstract] OR "facial beauty"[Title/Abstract] OR "facial profile"[Title/Abstract] OR "facial esthetics"[Title/Abstract] OR "facial aesthetics"[Title/Abstract] OR "facial esthetic"[Title/Abstract] OR "facial aesthetic"[Title/Abstract] OR "smile esthetics"[Title/Abstract] OR "smile aesthetics"[Title/Abstract] OR "smile attractiveness"[Title/Abstract] OR "smile design"[Title/Abstract] OR "dental esthetics"[Title/Abstract] OR "dental aesthetics"[Title/Abstract] OR "Esthetics, Dental"[MeSH Terms] OR "profile convexity"[Title/Abstract] OR "soft tissue profile"[Title/Abstract]) AND (orthodontic*[Title/Abstract] OR orthognathic[Title/Abstract] OR dental[Title/Abstract] OR dentistry[Title/Abstract] OR smile[Title/Abstract] OR malocclusion[Title/Abstract]) AND 2016:2026[dp]

4

Os registros duplicados foram identificados de forma automatizada pelo identificador PubMed (PMID) e pelo identificador de objeto digital (DOI) e removidos antes da triagem.

CRITÉRIOS DE ELEGIBILIDADE

Foram incluídos estudos primários que desenvolveram, validaram ou aplicaram modelos de IA, incluindo aprendizado de máquina, aprendizado profundo, ferramentas comerciais e grandes modelos de linguagem, para avaliar a estética da face ou do sorriso em contexto odontológico, ortodôntico ou bucomaxilofacial. Consideraram-se como análise estética a atribuição de escores de atratividade, a estimativa da idade aparente, a classificação estética do perfil facial, a avaliação de índices estéticos e a análise de pontos e proporções com finalidade estética. Foram aceitos estudos publicados entre 2016 e 2026, em português, inglês ou espanhol, com texto completo disponível gratuitamente, uma vez que a avaliação do relato metodológico exige a leitura integral dos métodos.

Foram excluídos revisões, editoriais, relatos de caso e resumos de eventos; estudos voltados à geração ou simulação de resultados, como o desenho automático do sorriso e a previsão do perfil pós-tratamento; estudos de diagnóstico ou de necessidade de tratamento sem finalidade estética; estudos sobre cor e translucidez dentárias; estudos em que a IA não participava da análise estética; e estudos fora do contexto odontológico.

SELEÇÃO DOS ESTUDOS E EXTRAÇÃO DOS DADOS

Os registros foram triados por título e resumo, e os estudos potencialmente elegíveis tiveram o texto completo buscado e avaliado quanto à elegibilidade. A triagem, a avaliação da elegibilidade e a extração dos dados foram realizadas pelos autores, com base no texto completo de todos os estudos incluídos.

Foram extraídos: país do primeiro autor, ano, tarefa estética, tipo de modelo e tamanho da amostra. Os desfechos metodológicos foram: (a) relato do sexo, (b) da idade e (c) da etnia, raça ou ancestralidade dos participantes, registrando-se se a amostra pertencia a uma única população ou a múltiplas; (d) relato da origem dos dados; (e) realização de validação externa, definida como a avaliação do modelo em dados de outro centro, população ou período não utilizados em seu desenvolvimento, aplicável aos modelos desenvolvidos pelos autores; (f) descrição da população de treinamento, aplicável às ferramentas prontas e aos modelos de linguagem; (g) disponibilidade dos dados; e (h) discussão do viés demográfico como limitação.

5

ANÁLISE DOS DADOS E ASPECTOS ÉTICOS

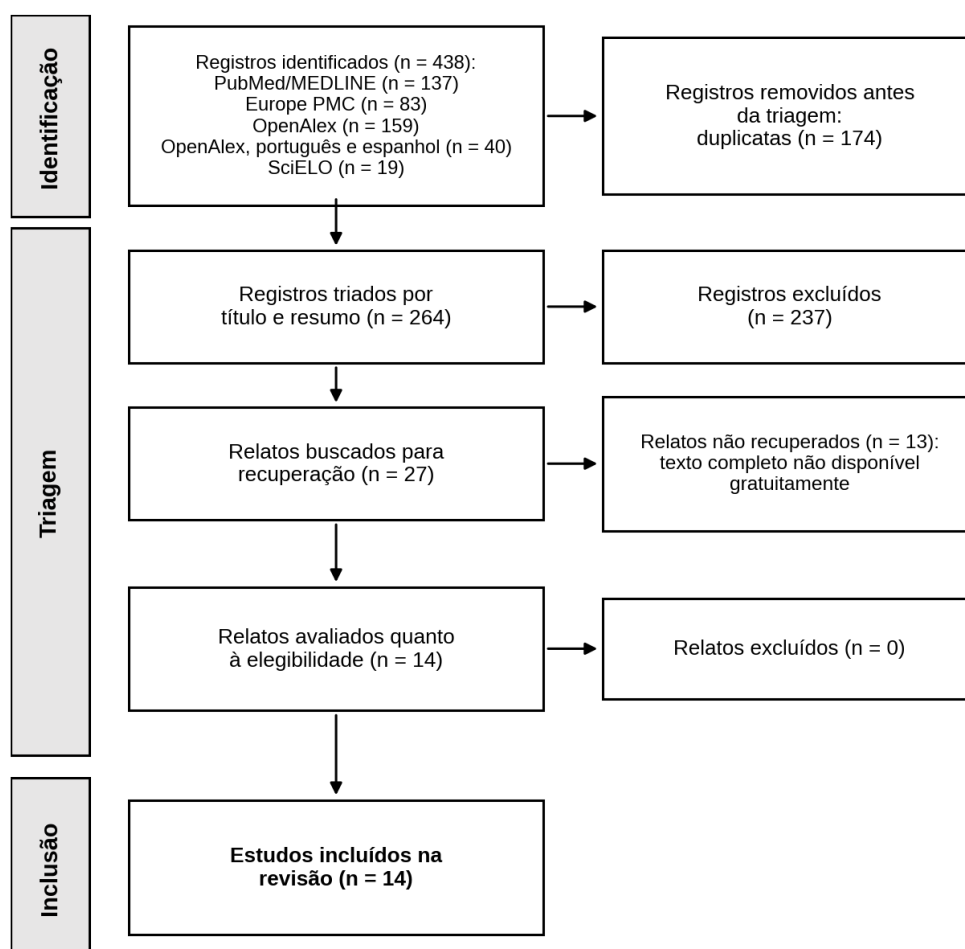
Os dados foram analisados de forma descritiva, por frequências absolutas e relativas. Por utilizar exclusivamente dados de estudos publicados, sem envolvimento direto de seres humanos, a pesquisa dispensou apreciação por Comitê de Ética em Pesquisa. Ferramentas de inteligência artificial generativa foram utilizadas como apoio na busca, na organização dos dados e na redação, e todo o conteúdo foi verificado pelos autores, que assumem integral responsabilidade pelo trabalho.

RESULTADOS

As buscas identificaram 438 registros: 137 no PubMed/MEDLINE, 83 no Europe PMC, 159 no OpenAlex, 40 na busca complementar em português e espanhol e 19 na SciELO. Após a remoção de 174 duplicatas, 264 registros foram triados por título e resumo, e 237 foram

excluídos. As buscas em português e espanhol e na SciELO não identificaram novos estudos elegíveis, pois os registros recuperados correspondiam a revisões, materiais educativos ou estudos fora do escopo odontológico. Dos 27 relatos buscados, 13 não foram recuperados por não terem texto completo disponível gratuitamente. Os 14 relatos avaliados quanto à elegibilidade atenderam aos critérios e foram incluídos (Figura 1).

Figura 1 - Fluxograma de seleção dos estudos, adaptado do PRISMA 2020.



Fonte: SOARES DP, et al., 2026; dados extraídos do PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO; adaptado de PAGE MJ, et al., 2021.

Os estudos foram publicados entre 2022 e 2026, com oito (57,1%) publicados em 2026, e vieram de 12 países, com maior número no Japão e na Turquia (n = 2 cada). A tarefa mais frequente foi a atribuição de escores de atratividade ou de estética facial (42,9%). Oito estudos (57,1%) desenvolveram modelos próprios, cinco (35,7%) aplicaram ferramentas prontas ou pré-treinadas e um (7,1%) utilizou grandes modelos de linguagem. Apenas um estudo tinha

primeiro autor vinculado a instituição brasileira, com análise de imagens de celebridades internacionais (Tabela 1).

Tabela 1 - Características dos estudos incluídos (n = 14).

Característica	n	%
Ano de publicação		
2022	2	14,3
2024	3	21,4
2025	1	7,1
2026	8	57,1
País do primeiro autor		
Japão	2	14,3
Turquia	2	14,3
Outros (10 países, 1 estudo cada)	10	71,4
Tarefa estética		
Escore de atratividade ou estética facial	6	42,9
Índices estéticos gengivais e dentários	2	14,3
Classificação do perfil facial	2	14,3
Características e expressão do sorriso	2	14,3
Análise de pontos e proporções	1	7,1
Idade facial percebida	1	7,1
Tipo de modelo		
Desenvolvido pelos autores	8	57,1
Ferramenta pronta ou pré-treinada	5	35,7
Grande modelo de linguagem	1	7,1

Fonte: SOARES DP, et al., 2026; dados extraídos do PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO.

O Quadro 1 apresenta os dados extraídos de cada estudo. Entre os que relataram etnia ou ancestralidade, três utilizaram amostras de uma única população: mulheres do leste asiático (TANIKAWA C, et al., 2024), pacientes chineses (BOONIPAT T, et al., 2025) e indivíduos caucasianos (PERDIKARIS PT, et al., 2026). Os dois estudos com amostras de múltiplas populações analisaram imagens públicas de celebridades de seis grupos étnicos (GRILLO R, et al., 2026) ou uma amostra planejada para refletir a composição racial dos Estados Unidos (STETZEL L, et al., 2024) (Quadro 1).

Quadro 1 - Relato de características demográficas, origem dos dados e validação nos estudos incluídos (n = 14).

Autor (ano)	País	Tipo	Amostra	Sexo	Idade	Etnia	Origem dos dados	Validação / treino
Lee S e Kim JE (2022)	Coreia do Sul	D	Base pública FFHQ	N	N	N	Base pública	Interna
Obwegeser D, et al. (2022)	Suíça	D	40 mulheres (960 fotos)	S	S	N	Centro único	Interna
Ma Q, et al. (2024)	Japão	D	133 pacientes	S	S	N	Centro único	Interna
Stetzel L, et al. (2024)	Estados Unidos	D	1.009 imagens	N	N	S (múltipla)	NR	Interna

Autor (ano)	País	Tipo	Amostra	Sexo	Idade	Etnia	Origem dos dados	Validação / treino
Tanikawa C, et al. (2024)	Japão	F	152 pacientes	S	S	S (única)	Multicêntrico	Treino descrito
Boonipat T, et al. (2025)	Taiwan	F	216 pacientes (864 fotos)	S	S	S (única)	Centro único	Treino NR
Azis H, et al. (2026)	Indonésia	D	482 imagens	N	N	N	Centro único	Interna
Bardakci E e Çelikel P (2026)	Turquia	LLM	9 variações de 1 imagem sintética	S	N	N	Imagem sintética	Treino NR
Elverişli Ö, et al. (2026)	Turquia	F	27 pacientes	S	S	N	NR	Treino NR
Grillo R, et al. (2026)	Brasil	F	156 celebridades	S	N	S (múltipla)	Imagens públicas	Treino NR
Ibrahim M, et al. (2026)	Egito	D	390 pacientes	S	S	N	Centro único	Interna
Perdikaris PT, et al. (2026)	Chipre	D	1.500 imagens (base SCUT-FBP ₅₅₀₀)	S	S	S (única)	Base pública	Interna
Pirayesh Z, et al. (2026)	Irã	D	400 pacientes	S	S	N	Centro único	Interna
Schirripa M (2026)	Itália	F	44 mulheres	S	N	N	Imagens públicas	Treino NR

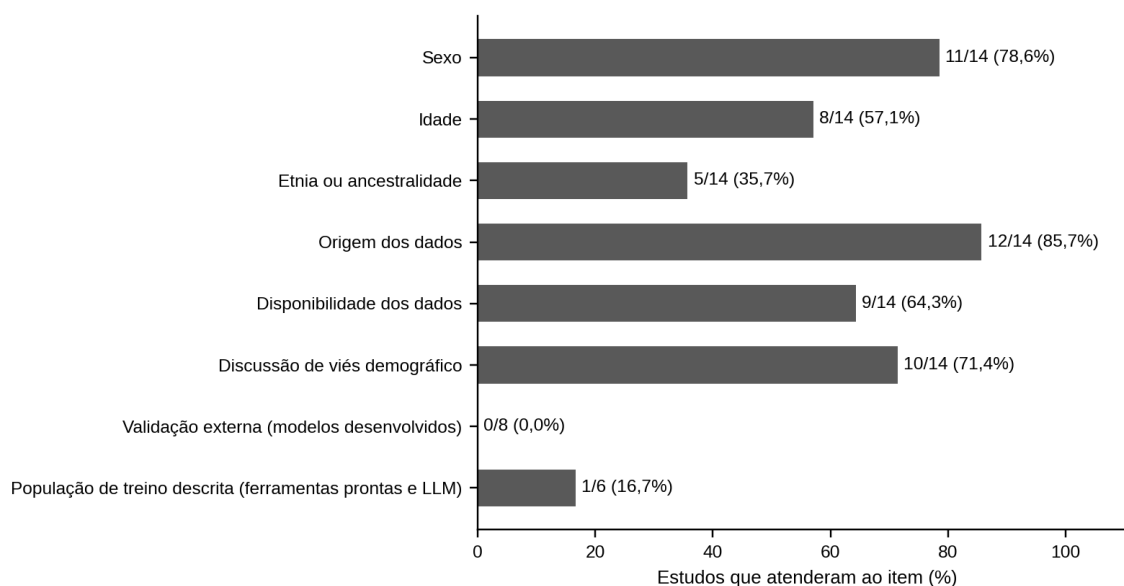
Legenda: Tipo: D, modelo desenvolvido pelos autores; F, ferramenta pronta ou pré-treinada; LLM, grande modelo de linguagem. S, relatado; N e NR, não relatado. Validação: aplicável aos modelos desenvolvidos; Treino: descrição da população de treinamento, aplicável às ferramentas prontas e aos modelos de linguagem.

Fonte: SOARES DP, et al., 2026; dados extraídos do PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO.

O sexo dos participantes foi relatado em 78,6% dos estudos, a idade em 57,1% e a etnia ou ancestralidade em 35,7%. A origem dos dados foi informada em 85,7% dos estudos, com predomínio de centro único (n = 6), e apenas um estudo utilizou dados de mais de um centro. Os dados estavam disponíveis em 64,3% dos estudos, sendo dois de forma pública e sete mediante solicitação, e 71,4% discutiram o viés demográfico como limitação (Gráfico 1).

8

Gráfico 1 - Proporção de estudos que atenderam a cada item de relato e de validação.



Fonte: SOARES DP, et al., 2026; dados extraídos do PubMed/MEDLINE, Europe PMC, OpenAlex e SciELO.

Nenhum dos oito modelos desenvolvidos pelos autores passou por validação externa, e todos foram avaliados apenas em dados internos. Em um estudo, os autores denominaram de externa a avaliação em fotografias provenientes do mesmo arquivo institucional usado no treinamento, o que, pelo critério desta revisão, corresponde a validação interna (AZIS H, et al., 2026). Entre as seis ferramentas prontas e modelos de linguagem aplicados, apenas um estudo (16,7%) descreveu a população em que a ferramenta havia sido treinada (TANIKAWA C, et al., 2024). Somente dois estudos analisaram o desempenho do modelo separadamente por sexo (PERDIKARIS PT, et al., 2026; PIRAYESH Z, et al., 2026), e nenhum o fez por grupo étnico. Nenhum estudo avaliou pacientes brasileiros.

DISCUSSÃO

Esta revisão mostra que os modelos de IA aplicados à análise estética facial e do sorriso vêm sendo desenvolvidos e testados em populações pouco descritas e pouco diversas, sem validação externa. Esses achados confirmam, para a área estética, os problemas apontados na IA ortodôntica de forma geral (ALKADHI OH, 2026) e indicam que a pergunta do título, sobre quem o algoritmo enxerga, ainda não pode ser respondida para a maioria dos modelos.

RELATO DEMOGRÁFICO INCOMPLETO

Mesmo com a leitura integral de todos os estudos, cerca de dois em cada dez não informaram o sexo dos participantes, quatro em cada dez não relataram a idade, e quase dois terços não descreveram a etnia ou ancestralidade da amostra. Um modelo, por exemplo, utilizou sexo, idade e peso como variáveis de entrada, mas não registrou a etnia dos 133 pacientes, todos atendidos em um único hospital (MA Q, et al., 2024). Em outro caso, a base de dados utilizada foi descrita como diversa em idade e etnia, mas sem a composição do subconjunto efetivamente usado no treinamento (LEE S e KIM JE, 2022).

Sem essas informações, o leitor não consegue julgar se um modelo é aplicável aos seus pacientes, o que contraria as recomendações de relato para modelos de aprendizado de máquina (COLLINS GS, et al., 2024). Descrever a composição demográfica da amostra é uma medida simples, que não exige recursos adicionais, mas que aumenta de forma importante a transparência dos estudos.

AMOSTRAS DE POPULAÇÃO ÚNICA E PADRÕES ESTÉTICOS

Entre os estudos que relataram etnia, a maioria utilizou amostras de uma única população. Um modelo de atratividade facial, por exemplo, foi treinado exclusivamente com imagens de indivíduos caucasianos, e os próprios autores observaram que o erro foi menor no segmento mais densamente representado da base de dados (PERDIKARIS PT, et al., 2026). Esse resultado ilustra um princípio geral: o modelo acerta mais onde viu mais exemplos, e tende a errar mais nos grupos menos representados.

O estudo que comparou grupos étnicos reforça a importância dessa questão. A análise de celebridades de seis grupos identificou padrões comuns, mas também variações entre grupos em proporções, simetria e variabilidade morfológica, e as imagens foram obtidas em mídias e redes sociais a partir de buscas por termos associados à beleza, o que tende a reproduzir os padrões estéticos veiculados pela própria mídia (GRILLO R, et al., 2026). Além disso, a avaliação da atratividade do perfil variou conforme o sexo do avaliador e do indivíduo avaliado (PIRAYESH Z, et al., 2026). Um modelo treinado com uma única população pode, portanto, transformar o padrão estético dessa população em referência para todos, como já se observou com algoritmos em outras áreas da saúde (OBERMEYER Z, et al., 2019).

10

FERRAMENTAS PRONTAS E MODELOS DE LINGUAGEM: A POPULAÇÃO INVISÍVEL

Quase todos os estudos que aplicaram ferramentas prontas não descreveram a população em que essas ferramentas foram treinadas. Um detector automático de pontos faciais foi usado para definir padrões de atratividade em diferentes grupos étnicos sem informação sobre seus dados de treinamento (GRILLO R, et al., 2026); um *software* comercial de análise de expressões foi aplicado a pacientes chineses (BOONIPAT T, et al., 2025); um sistema de análise do sorriso foi aplicado a imagens públicas de vencedoras de concursos de beleza (SCHIRRIPA M, 2026); e um aplicativo de estimativa de idade foi utilizado sem validação publicada, como reconhecido pelos próprios autores (ELVERIŞLI Ö, et al., 2026).

O contraste aparece em um estudo que descreveu sua base de treinamento: a rede de atratividade aprendeu com imagens de um aplicativo de relacionamentos, compostas majoritariamente por caucasianos, e os autores ressaltaram que o modelo reflete apenas o que os usuários dessa base consideram atraente (OBWEGESER D, et al., 2022). O mesmo problema ocorre com os grandes modelos de linguagem, cujos dados de treinamento não são públicos. No

estudo que os avaliou, a imagem analisada era um sorriso infantil gerado por IA, e os modelos atribuíram notas mais altas e homogêneas que os avaliadores humanos, inclusive a imagens com lesões de cárie (BARDAKCI E e ÇELIKEL P, 2026). Quando uma ferramenta desse tipo é usada como instrumento de medida, seus vieses passam a fazer parte dos resultados do estudo.

AUSÊNCIA DE VALIDAÇÃO EXTERNA

Nenhum modelo desenvolvido pelos autores foi avaliado em dados de outro centro ou população. A validação interna, mesmo com conjuntos de teste separados, avalia o desempenho em pacientes semelhantes aos do treinamento e tende a superestimar a acurácia na prática clínica, como mostram as perdas de desempenho observadas em testes entre instituições (ALKADHI OH, 2026). O uso impreciso do termo validação externa para testes realizados com dados da mesma instituição (AZIS H, et al., 2026) indica que o próprio conceito ainda precisa ser melhor difundido na área.

Por outro lado, a disponibilidade de dados foi relativamente frequente, ainda que majoritariamente mediante solicitação. Apenas dois estudos disponibilizaram publicamente a base de dados ou o código e os modelos treinados, o que permite a replicação e a validação independente por outros pesquisadores (OBWEGESER D, et al., 2022; PERDIKARIS PT, et al., 2026).

IMPLICAÇÕES PARA O BRASIL E RECOMENDAÇÕES

Nenhum estudo avaliou pacientes brasileiros, e a única publicação com primeiro autor brasileiro analisou imagens de celebridades internacionais (GRILLO R, et al., 2026). Considerando que a maior parte da população brasileira se declara parda ou preta (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2023), ferramentas treinadas predominantemente com populações caucasianas ou asiáticas podem ter desempenho inferior no país, sem que isso seja percebido pelo clínico.

Recomenda-se que futuros estudos relatem sexo, idade e cor ou raça dos participantes, preferencialmente segundo as categorias do IBGE no caso de amostras brasileiras; utilizem dados de mais de um centro; realizem validação externa; apresentem o desempenho por subgrupos demográficos; e disponibilizem dados ou códigos sempre que possível. Aos clínicos, recomenda-se verificar em que população uma ferramenta foi desenvolvida antes de utilizá-la no planejamento estético.

LIMITAÇÕES DO ESTUDO

Esta revisão apresenta limitações. A inclusão restrita a estudos com texto completo disponível gratuitamente pode ter introduzido viés de seleção, uma vez que 13 relatos potencialmente elegíveis não puderam ser avaliados. A seleção e a extração não foram realizadas de forma independente por dois revisores, a busca não incluiu as bases Scopus, Web of Science, Embase e LILACS, e o protocolo não foi registrado previamente. Por fim, o escopo restringiu-se aos modelos de análise estética, de modo que os achados não se estendem necessariamente aos modelos de geração ou previsão de resultados.

CONSIDERAÇÕES FINAIS

Os modelos de IA aplicados à análise estética facial e do sorriso são, em sua maioria, desenvolvidos e testados em populações pouco descritas e pouco diversas. Quase dois terços dos estudos não relataram a etnia ou ancestralidade dos participantes, nenhum modelo desenvolvido pelos autores passou por validação externa e quase nenhuma ferramenta pronta teve sua população de treinamento descrita.

Esses achados limitam a generalização dessas ferramentas para populações miscigenadas, como a brasileira, e indicam que seu uso clínico deve ser cauteloso. O desenvolvimento de bases de dados multicêntricas e diversas, o relato padronizado das características demográficas e a validação externa são passos necessários para que a IA contribua para uma estética mais objetiva sem reproduzir padrões de beleza restritos a determinadas populações.

REFERÊNCIAS

1. ALKADHI OH. Artificial intelligence in orthodontics: from laboratory benchmarks to clinical care. *Cureus*, 2026; 18(6): e111151. DOI: 10.7759/cureus.111151.
2. AZIS H, et al. Deep learning approaches for soft tissue classification of cephalometric images in orthodontics. *JOIV: International Journal on Informatics Visualization*, 2026; 10(3): 988. DOI: 10.62527/joiv.10.3.3885.
3. BARDAKCI E, ÇELIKEL P. Artificial intelligence versus human evaluation of pediatric smile aesthetics: a comparative study across multiple observer groups. *Journal of Health Sciences and Medicine*, 2026; 9(3): 830-838. DOI: 10.32322/jhsm.1920254.

4. BOONIPAT T, et al. Impact of orthognathic surgery on emotional expression and smile aesthetics in Chinese patients as measured by FaceReader: a before-and-after study. *Archives of Craniofacial Surgery*, 2025; 26(3): 115-123. DOI: 10.7181/acfs.2025.0007.
5. COLLINS GS, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 2024; 385: e078378. DOI: 10.1136/bmj-2023-078378.
6. ELVERİŞLİ Ö, et al. Assessment of perceived facial age changes following orthognathic surgery using artificial intelligence. *Healthcare*, 2026; 14(14): 2200. DOI: 10.3390/healthcare14142200.
7. GRILLO R, et al. Frontal facial analysis of male celebrities attractiveness standards through artificial intelligence. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 2026; 127(4): 102776. DOI: 10.1016/j.jormas.2026.102776.
8. IBRAHIM M, et al. Artificial intelligence-based photographic detection of pink esthetic score attributes using a hybrid deep learning segmentation pipeline: a method development study. *Scientific Reports*, 2026; 16(1): 23726. DOI: 10.1038/s41598-026-62229-4.
9. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Demográfico 2022: identificação étnico-racial da população, por sexo e idade: resultados do universo. Rio de Janeiro: IBGE, 2023.
10. LEE S, KIM JE. Evaluating the precision of automatic segmentation of teeth, gingiva and facial landmarks for 2D digital smile design using real-time instance segmentation network. *Journal of Clinical Medicine*, 2022; 11(3): 852. DOI: 10.3390/jcm11030852.
11. MA Q, et al. 3D evaluation model of facial aesthetics based on multi-input 3D convolution neural networks for orthognathic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2024; 20(3): e2651. DOI: 10.1002/rcs.2651.
12. MURAD MH, WANG Z. Guidelines for reporting meta-epidemiological methodology research. *Evidence-Based Medicine*, 2017; 22(4): 139-142. DOI: 10.1136/ebmed-2017-110713.
13. OBERMEYER Z, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 2019; 366(6464): 447-453. DOI: 10.1126/science.aax2342.
14. OBWEGESER D, et al. Using artificial intelligence to determine the influence of dental aesthetics on facial attractiveness in comparison to other facial modifications. *European Journal of Orthodontics*, 2022; 44(4): 445-451. DOI: 10.1093/ejo/cjaco16.
15. PAGE MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 2021; 372: n71. DOI: 10.1136/bmj.n71.
16. PERDIKARIS PT, et al. Artificial intelligence for facial attractiveness assessment in orthodontics: model development and evaluation. *European Journal of Orthodontics*, 2026; 48(5): cjado68. DOI: 10.1093/ejo/cjado68.

17. PIRAYESH Z, et al. Predicting perceived profile attractiveness from cephalometric measurements using machine learning. *International Dental Journal*, 2026; 76(4): 109679. DOI: 10.1016/j.identj.2026.109679.
18. SCHIRRIPA M. Smile evolution in the social media era: a longitudinal analysis of Miss Italia winners (1982-2025). *Annali di Stomatologia*, 2026; 17(2): 507-516. DOI: 10.59987/ads/2026.2.507-516.
19. STETZEL L, et al. Artificial intelligence for predicting the aesthetic component of the Index of Orthodontic Treatment Need. *Bioengineering*, 2024; 11(9): 861. DOI: 10.3390/bioengineering11090861.
20. TANIKAWA C, et al. Facial soft-tissue shape changes after fixed edgewise treatment with premolar extraction in individual artificial-intelligence-classified facial profile patterns. *BMC Oral Health*, 2024; 24(1): 740. DOI: 10.1186/s12903-024-04512-2.