Estimación fiable de la edad cronológica a partir del procesamiento automático de imágenes panorámicas dentales

A. Vivel-Couso, N. Vila-Blanco, M.J. Carreira, A. Bugarin-Diz, I. Tomás, and J.M. Alonso-Moral

Centro Singular de Investigacion en Tecnoloxías Intelixentes (CiTIUS)
Universidade de Santiago de Compostela
Santiago de Compostela, Spain
{ainhoa.vivel.couso, nicolas.vila, mariajose.carreira, alberto.bugarin.diz, inmaculada.tomas, josemaria.alonso.moral}@usc.es

Resumen En este trabajo, presentamos un resumen del artículo titulado "Trustworthy Data-driven Chronological Age Estimation from Panoramic Dental Images", de A. Vivel-Couso et al. [4], publicado en la revista *Information Systems Frontiers*. El artículo fue enviado el 21 de julio de 2025 y publicado el 15 de enero de 2026.

Keywords: Inteligencia Artificial Explicable centrada en las Personas · Generación de Texto a partir de Datos · Generación de Texto basada en Reglas · Cuantificación Lingüística basada en Lógica Borrosa

1. Introducción

En este artículo presentamos un resumen de nuestro trabajo titulado *"Trustworthy Data-driven Chronological Age Estimation from Panoramic Dental Images"* [4], publicado en la revista *Information Systems Frontiers*, donde describimos el sistema AgeX, que genera explicaciones textuales sobre la estimación de la edad cronológica de una persona a partir del procesamiento automático de imágenes panorámicas dentales mediante un modelo de aprendizaje neuronal profundo no transparente. La estimación de la edad cronológica es importante en aplicaciones como la identificación forense, los procedimientos legales, el control migratorio y los procesos de adopción, entre otros.

AgeX sigue un enfoque de fiabilidad desde el diseño (*trustworthiness-by-design*), incorporando requisitos clínicos de interpretabilidad, robustez y explicabilidad [1]. En la práctica, es un sistema inteligente que combina componentes opacos e interpretables para proporcionar tanto predicciones precisas como explicaciones comprensibles. Se alinea con los requisitos del Reglamento de IA de la UE (EU AI Act[1]) al mejorar la transparencia mediante explicaciones automáticas en lenguaje natural e informar a los usuarios cuando se utiliza contenido generado por IA. La fiabilidad y la utilidad del enfoque fueron evaluadas por profesionales dentales expertos en un escenario de despliegue realista.

[1] https://artificialintelligenceact.eu

2. Propuesta

AgeX incluye un módulo de Generación de Lenguaje Natural (NLG) para construir explicaciones textuales fiables, siguiendo el "pipeline" clásico basado en reglas [2,3], que consta de tres etapas: (i) planificación del documento, (ii) micro-planificación, y (iii) realización lingüística.

En la etapa de planificación del documento, se definen el contenido narrativo y su estructura según un conjunto de reglas derivadas de informes redactados por expertos, que se toman como ejemplos de referencia, los cuales fueron validados por un especialista en odontología y codificados en plantillas reutilizables.

Durante la etapa de microplanificación, el sistema determina cómo se expresa la información. Un mecanismo basado en reglas distingue entre oraciones estáticas (explicaciones generales sobre el método de estimación considerado e independientes del caso particular bajo estudio) y oraciones dinámicas (contenido específico del paciente derivado de datos clínicos). A continuación, un conjunto de reglas condicionales determina qué información se incluye en la narrativa a generar. Además, se utilizan cuantificadores borrosos [5] (por ejemplo, *ninguno, pocos, algunos, la mayoría, todos...*) para transformar datos dentales numéricos en descripciones lingüísticas comprensibles para las personas.

En la etapa de realización lingüística, el uso de `SimpleNLG-ES` garantiza la correcta generación sintáctica, morfológica y gramatical de las narrativas finales.

El sistema se desarrolló de forma iterativa, con supervisión continua de expertos en odontología en cada etapa. Se aplicó un enfoque de generación de texto a partir de datos (Data-to-Text) basado en reglas en lugar de métodos de NLG neuronales, lo que facilita la supervisión humana, garantiza controlabilidad y fiabilidad desde el diseño. La validación final por expertos del dominio se utilizó como criterio principal de evaluación, en línea con las mejores prácticas para sistemas de NLG basados en reglas.

3. Discusión de Resultados

Se utilizó el módulo de NLG de AgeX para generar informes textuales correspondientes a cuatro casos clínicos de pacientes de entre 5 y 16 años. Cada caso incluía la edad cronológica real y la estimada por el modelo opaco. Siete expertos en odontología evaluaron, en una escala de 1 a 5, los cuatro informes generados por el sistema mediante un cuestionario que incluía 5 afirmaciones a valorar. Los resultados resumidos en la Tabla 1 muestran los distintos niveles de calidad percibida por los expertos en los tres escenarios considerados.

En el primer escenario (E1), únicamente se proporcionó información contextual mínima, junto con la estimación de la edad. Esta configuración obtuvo la puntuación media más baja (3,94). Aunque la terminología (4,57), la comprensibilidad (4,71) y la coherencia con la radiografía panorámica (4,43) fueron valoradas positivamente, los expertos señalaron como principales debilidades la insuficiencia de información (2,54) y la utilidad limitada (3,42). Este escenario también presentó la mayor variabilidad entre los expertos ($\sigma = 0{,}83$), y solo el 3,57 % lo seleccionó como la mejor opción.

Tabla 1. Media ($\bar{x}$) y desviación estándar (σ) de las puntuaciones otorgadas por los expertos en odontología. Se consideran tres escenarios: Escenario 1 (E1), los evaluadores disponen únicamente de la estimación de la edad dental; Escenario 2 (E2), la estimación se acompaña de una breve descripción automática; Escenario 3 (E3), la estimación se acompaña de una descripción automática completa y detallada.

Aspectos a valorar	**E1**	**E2**	**E3**
1. **Terminología**: "La descripción textual utiliza la terminología adecuada para describir la dentición de los pacientes."	4,57	4,54	4,64
2. **Comprensibilidad**: "La descripción es fácil de entender."	4,71	4,69	4,60
3. **Completitud**: "La descripción textual proporcionada contiene información suficiente."	2,54	3,50	4,89
4. **Utilidad**: "La descripción textual proporcionada contiene información útil."	3,42	4,07	4,82
5. **Coherencia**: "La descripción textual es coherente con respecto a la imagen (radiografía panorámica o mapa de calor) que acompaña."	4,43	4,61	4,89
Media ($\bar{x}$)	3,94	4,28	4,77
Desviación estándar (σ)	0,83	0,44	0,12

En el segundo escenario (E2) se añadió una descripción breve junto a la radiografía panorámica, lo que resultó en una mejora de la evaluación media ($\bar{x} = 4{,}28$). Los expertos consideraron que la calidad narrativa (en términos de terminología, comprensibilidad y coherencia) era buena ($\bar{x} > 4$), aunque la completitud de la información seguía siendo limitada (3,5). La variabilidad disminuyó ($\sigma = 0{,}44$), pero ninguno de los evaluadores seleccionó este escenario como el mejor.

En el tercer escenario (E3) se proporcionó a los expertos una explicación detallada junto con el correspondiente mapa de calor asociado a la imagen. Esta configuración obtuvo la evaluación media más alta ($\bar{x} = 4{,}77$). Los evaluadores calificaron de forma consistentemente positiva el sistema, destacando una terminología adecuada (4,64), una alta comprensibilidad (4,60), completitud (4,89) y utilidad (4,82) de la información. Las explicaciones generadas también se consideraron altamente coherentes con las imágenes correspondientes (4,89). Este escenario presentó la menor variabilidad entre los evaluadores ($\sigma = 0{,}12$), lo que indica un fuerte consenso y una menor incertidumbre en comparación con los escenarios anteriores. Además, el 89,29 % de los expertos seleccionó esta configuración como la mejor opción. Estos resultados indican una clara preferencia por las explicaciones completas generadas por el módulo NLG de AgeX, lo que confirma su eficacia como narrativas fiables e informativas de apoyo a la toma de decisiones clínicas.

Adicionalmente, la fiabilidad del sistema AgeX se evaluó considerando las 7 dimensiones de la lista de verificación ALTAI[2]. Las dimensiones con menor puntuación fueron *Agencia humana y supervisión* (3,66) y *Bienestar ambiental y social* (3,75).

[2] https://altai.insight-centre.org

En relación con *Agencia humana y supervisión*, AgeX no realiza toma de decisiones autónoma. El sistema funciona como una herramienta de apoyo para expertos en odontología, quienes mantienen la plena responsabilidad de las decisiones clínicas. Para reforzar este aspecto, se añadió una cláusula de exención de responsabilidad para enfatizar explícitamente la autonomía humana y la supervisión en el uso del sistema.

En cuanto a *Bienestar ambiental y social*, las consideraciones se extienden al impacto más amplio de los sistemas de IA, incluyendo la huella ambiental y la sostenibilidad. Aunque AgeX se basa en un módulo de NLG basado en reglas y en un modelo predictivo ligero, lo que resulta en un impacto ambiental prácticamente despreciable, se requiere mayor atención a la formación de usuarios y al desarrollo de competencias para maximizar el beneficio social a largo plazo.

Agradecimientos

Ainhoa Vivel Couso es una investigadora predoctoral financiada por la Consellería de Educación, Ciencia, Universidades e Formación Profesional de Galicia (ED481A-2024-059). Los autores reconocen el apoyo de la Consellería de Educación, Ciencia, Universidades e Formación Profesional de Galicia (proyectos CiTIUS 2024-2027 ED431G2023/04, GSI 2022-2025 ED431C2022/19, GVA GRC2021/48 and OSRG GPC2020/27) y de la Unión Europea (Fondo Europeo de Desarrollo Regional - FEDER). Este trabajo también está financiado por el Ministerio de Ciencia, Innovación y Universidades de España (MCIN/AEI/10.13039/501100011033/) a través de los proyectos XAI4SOC PID2021-123152OB-C21, MAIXAI4Trust PID2024-157680NB-I00, y COLLAGE PID2023-149549NB-I00, así como por el Instituto de Salud Carlos III mediante PI21/00588.

Referencias

1. Lekadir, K., Quaglio, G., Garmendia, A.T., Gallin, C.: Artificial intelligence in healthcare: applications, risks, and ethical and social impacts (2022). https://doi.org/10.2861/568473
2. Reiter, E.: Natural Language Generation. Springer Cham (2024). https://doi.org/10.1007/978-3-031-68582-8
3. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Studies in Natural Language Processing, Cambridge University Press (2000)
4. Vivel-Couso, A., Vila-Blanco, N., Carreira, M.J., Bugarín-Diz, A., Tomás, I., Alonso-Moral, J.M.: Trustworthy data-driven chronological age estimation from panoramic dental images. Information Systems Frontiers (2026). https://doi.org/10.1007/s10796-025-10682-3
5. Zadeh, L.: A computational approach to fuzzy quantifiers in natural languages. Computers & Mathematics with Applications **9**(1), 149–184 (1983). https://doi.org/10.1016/0898-1221(83)90013-5