9.5. Inteligencia artificial en la diabetes tipo 2

PUNTOS CLAVE

1. La única aplicación de la IA en la DM2 que hoy está autorizada y funcionando de forma autónoma es el cribado de la retinopatía diabética sobre retinografía.
2. Los modelos de predicción de progresión renal superan claramente a la estratificación clínica convencional, pero todavía no están implantados en España.
3. El resto de aplicaciones —predicción de riesgo, asistentes conversacionales, apoyo a la decisión— tienen evidencia preliminar y poca implantación real.
4. El riesgo principal no es que la IA se equivoque, sino que se equivoque de forma sistemática en los grupos peor representados en los datos con los que se entrenó.
5. La IA no sustituye el juicio clínico: la decisión sigue exigiendo contexto, valoración biopsicosocial y acuerdo con la persona.

La inteligencia artificial engloba tecnologías —aprendizaje automático, redes neuronales profundas, procesamiento del lenguaje natural— capaces de analizar grandes volúmenes de datos y detectar patrones difíciles de identificar con los métodos estadísticos habituales (Kavakiotis I, 2017; Topol EJ, 2019). En la DM2 su potencial es considerable, pero conviene separar lo que ya funciona de lo que todavía es promesa: la distancia entre ambas cosas es el contenido útil de este capítulo (Tabla 1).

Tabla 1. Aplicaciones de la inteligencia artificial en la DM2: qué promete y dónde está

Ámbito	Qué promete	Dónde está hoy
Cribado de retinopatía diabética	Lectura automatizada de la retinografía sin intervención del oftalmólogo.	Es la aplicación más madura. Un sistema autónomo evaluado en 900 personas atendidas en centros de Atención Primaria alcanzó una sensibilidad del 87,2 % y una especificidad del 90,7 %, y fue el primero autorizado por la FDA en cualquier especialidad médica (Abràmoff MD, 2018). Capítulo 7.6.
Predicción de progresión renal	Identificar a quien va a perder función renal antes de que ocurra.	Un modelo que combina biomarcadores e historia clínica electrónica en 1146 personas con enfermedad renal diabética obtuvo un AUC de 0,77 frente a 0,62 del modelo clínico convencional, con un valor predictivo positivo del 61 % en el grupo de alto riesgo frente al 40 % de la clasificación KDIGO (Chan L, 2021). Aún no incorporado a la práctica habitual en España. Capítulo 7.3.
Predicción de riesgo metabólico y cardiovascular	Integrar variables clínicas, analíticas y sociodemográficas para anticipar diabetes, insuficiencia cardíaca o eventos cardiovasculares.	Numerosos modelos publicados, con rendimiento superior a algunas escalas clásicas, pero con escasa validación externa y muy poca implantación real (Kavakiotis I, 2017).
Monitorización glucémica	Detectar patrones, anticipar tendencias y sugerir ajustes a partir de los datos del sensor.	Es donde la IA ya está funcionando de forma cotidiana, sobre todo en los sistemas automatizados de administración de insulina, de uso mayoritario en la DM1. En la DM2 el interés crece con la expansión de la monitorización continua. Capítulo 9.3.

Ámbito	Qué promete	Dónde está hoy
Asistentes conversacionales	Recordatorios, educación, apoyo motivacional y resolución de dudas frecuentes.	Línea emergente con evidencia todavía preliminar: los estudios disponibles son de viabilidad, con muestras pequeñas y en poblaciones distintas de la DM2 del adulto (Stephens TN, 2019).
Apoyo a la decisión clínica	Alertar sobre riesgo de hipoglucemia, interacciones, deterioro renal o inercia terapéutica.	Prometedor en un contexto de creciente complejidad terapéutica, pero con poca evidencia de impacto en resultados de salud (Contreras I, 2018).
Modelos de lenguaje generales	Responder preguntas clínicas complejas y apoyar decisiones terapéuticas en escenarios de incertidumbre.	Rendimiento distinto e inferior al del clínico: en un estudio comparativo, GPT-4 eligió metformina en el 12 % de los casos frente al 31 % de los endocrinólogos ante el mismo escenario de incertidumbre. No se recomienda su uso sin supervisión clínica (Flory JH, 2025).

AUC: área bajo la curva ROC; DM1: diabetes mellitus tipo 1; DM2: diabetes mellitus tipo 2; FDA: Food and Drug Administration; IA: inteligencia artificial; KDIGO: Kidney Disease Improving Global Outcomes.

Fuente: Elaboración propia a partir de las referencias citadas en cada fila.

1. Lo que ya está funcionando

El cribado automatizado de la retinopatía diabética es el ejemplo más sólido y, además, se validó exactamente en el escenario de esta guía. El ensayo pivotal incluyó a 900 personas con diabetes y sin retinopatía conocida atendidas en centros de Atención Primaria, comparando el sistema autónomo con la lectura estereoscópica de referencia: sensibilidad del 87,2 % (intervalo de confianza [IC] del 95 % 81,8-91,2), especificidad del 90,7 % (IC 95 % 88,3-92,7) y capacidad de obtener imágenes válidas en el 96,1 % de los casos. Sobre esos resultados la agencia estadounidense del medicamento (FDA) autorizó el primer sistema diagnóstico autónomo basado en inteligencia artificial (IA) de cualquier especialidad médica (Abràmoff MD, 2018). El interés para la Atención Primaria es evidente en áreas con dificultad de acceso a oftalmología. En España la implantación es desigual: Canarias cuenta con cribado sistemático en Atención Primaria desde 2009, mientras que en el resto del territorio es puntual; conviene comprobarlo por comunidad autónoma, porque cambia con rapidez.

En enfermedad renal diabética, un modelo que combina biomarcadores plasmáticos con datos de la historia clínica electrónica, evaluado en 1146 personas, alcanzó un área bajo la curva (AUC) de 0,77 tanto en derivación como en validación, frente al 0,62 del modelo clínico convencional. Entre los clasificados como de alto riesgo, el 61 % progresó realmente, frente al 40 % del estrato de mayor riesgo de la clasificación Kidney Disease: Improving Global Outcomes (KDIGO); y solo el 10 % de los clasificados como de bajo riesgo progresó (Chan L, 2021). Es una mejora real sobre la práctica actual, aunque todavía no esté disponible en España.

2. Lo que todavía no

En el resto de ámbitos la situación es más modesta de lo que sugiere el ruido mediático. Existen numerosos modelos predictivos de riesgo cardiovascular y metabólico con buen rendimiento en los datos con los que se construyeron, pero con escasa validación externa. Los asistentes conversacionales cuentan con estudios de viabilidad en muestras pequeñas y poblaciones que no son la DM2 del adulto. Y los sistemas de apoyo a la decisión clínica son prometedores en un contexto de complejidad terapéutica creciente, pero apenas hay evidencia de que mejoren resultados de salud (Contreras I, 2018).

3. Limitaciones y aspectos éticos

El problema más serio no es la exactitud media de los modelos, sino su comportamiento desigual. Los algoritmos aprenden de datos recogidos históricamente, de modo que los grupos que ya han sufrido sesgos —por etnia, nivel socioeconómico o edad— son precisamente los más vulnerables a una predicción incorrecta o a que se les retiren recursos. Evitarlo exige incorporar criterios de equidad al diseño, al despliegue y a la evaluación de los modelos, no solo confiar en que la exactitud global sea buena (Rajkomar A, 2018). El Capítulo 8.4 desarrolla por qué esto importa especialmente en la DM2.

A ello se añaden la opacidad de algunos algoritmos, que dificulta entender cómo se genera una recomendación; la privacidad y la protección de datos clínicos sensibles; la responsabilidad legal cuando la recomendación falla; y la posibilidad de que la incorporación indiscriminada de herramientas aumente la carga de trabajo en lugar de reducirla, si no se acompaña de formación e integración en los sistemas de información. A esto se suma el Reglamento (UE) 2024/1689 de IA, que añadirá requisitos de alto riesgo a sistemas como el cribado de retinopatía; su calendario se ha retrasado y conviene comprobarlo antes de publicar.

La Atención Primaria tiene aquí un papel que no es menor. Por su visión longitudinal, el equipo de Atención Primaria está en la mejor posición para seleccionar qué herramientas son útiles y seguras, interpretar críticamente lo que generan y evitar que una alerta automática se convierta en una cascada diagnóstica innecesaria. La IA es apoyo, no sustituto: la decisión terapéutica sigue exigiendo contextualización individual y acuerdo con la persona (ADA, 2026, secc. 7).

PARA LLEVARSE A LA CONSULTA

1. Si el área ofrece lectura automatizada de retinografías, conocerla: es la única aplicación de la IA en diabetes con autorización regulatoria y validación en atención primaria.
2. Ante cualquier modelo predictivo, preguntar en qué población se entrenó y en cuál se validó antes de aplicarlo en la consulta.
3. Desconfíe de las herramientas que no permiten saber en qué se basa su recomendación.
4. Tratar las alertas automáticas como una pregunta, no como una orden: la decisión sigue necesitando contexto clínico y acuerdo con la persona.
5. Vigilar el efecto sobre la equidad: un algoritmo que funciona peor en los pacientes más desfavorecidos amplía la brecha en lugar de reducirla.

Bibliografía

1. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39. doi:10.1038/s41746-018-0040-6
2. American Diabetes Association Professional Practice Committee for Diabetes. 7. Diabetes technology: Standards of Care in Diabetes—2026. Diabetes Care. 2026;49(Suppl 1):S150-65. doi:10.2337/dc26-S007
3. Chan L, Nadkarni GN, Fleming F, McCullough JR, Connolly P, Mosoyan G, et al. Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease. Diabetologia. 2021;64(7):1504-15. doi:10.1007/s00125-021-05444-0
4. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res. 2018;20(5):e10775. doi:10.2196/10775
5. Flory JH, Ancker JS, Kim SYH, Kuperman G, Petrov A, Vickers A. Large language model GPT-4 compared to endocrinologist responses on initial choice of glucose-lowering medication under conditions of clinical uncertainty. Diabetes Care. 2025;48(2):185-192. doi:10.2337/dc24-1067
6. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J. 2017;15:104-16. doi:10.1016/j.csbj.2016.12.005
7. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. doi:10.7326/M18-1990
8. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med. 2019;9(3):440-7. doi:10.1093/tbm/ibz043
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7