

DETECCIÓN DEL AUTISMO BASADA EN EL ANÁLISIS COMPUTACIONAL DEL LENGUAJE DE LOS PADRES

MARIA ELEONORA MINISSI, ALBERTO ALTOZANO, LUCIA GÓMEZ-ZARAGOZÁ, LUNA MADDALON, MARIANO ALCAÑIZ, JAVIER MARÍN-MORALES

Laboratorio de Neurotecnologías Inmersivas. Instituto Human-Tech.
Universidad Politécnica de Valencia. España

Dirección postal: Maria Eleonora Minissi, Instituto Human-Tech, Universidad Politécnica de Valencia. Camí de Vera s/n, 46022, Valencia. España

E-mail: meminiss@htech.upv.es

Resumen

La detección temprana del trastorno del espectro autista (TEA) exige integrar la observación clínica con la historia del desarrollo y el funcionamiento cotidiano, pero los circuitos asistenciales suelen verse tensionados por la alta demanda y los recursos especializados limitados. En este escenario, la narrativa de madres y padres aporta información ecológica sobre comunicación social, lenguaje y conductas repetitivas; sin embargo, dichos relatos son difíciles de analizar de forma sistemática en la práctica. Este artículo revisa un artículo que analizó automáticamente las respuestas abiertas de cuidadores a 12 preguntas inspiradas en el ADI-R, en 51 familias (niños 2-8 años; 25 TEA y 26 controles). El objetivo del estudio fue detectar de manera automática, a través del análisis computacional del habla de los padres, la presencia del TEA en los menores. La mejor estrategia a nivel de clasificación de sujeto se obtuvo transformando el texto en representaciones de significado (a través de *text-ada-embedding-large-v3* de OpenAI), entrenando los modelos por pregunta, incorporando la pregunta en la entrada y agregando las decisiones por "votación" (84% de precisión; ROC-AUC 1.0). El objetivo clínico del estudio piloto ha sido demostrar la posibilidad de apoyar el cribado, la priorización, la derivación y, potencialmente, reducir tiempos de espera y carga profesional, favoreciendo acceso oportuno a evaluación e intervención tempranas. Se discuten los límites (muestra pequeña, población seleccionada, influencia posible de intervención previa) y los riesgos éticos (sesgos sociolingüísticos, privacidad, estigmatización,

falsos positivos/negativos), y se proponen pasos para una traslación responsable.

Palabras clave: trastorno del espectro autista, detección temprana, narrativa parental, análisis computacional.

Abstract

Autism detection based on computational analysis of parental language

Early detection of autism spectrum disorder (ASD) requires integrating clinical observation with developmental history and everyday functioning, yet care pathways are often strained by high demand and limited specialist resources. In this context, parents' narratives provide ecologically valid signals about social communication, language, and repetitive behaviors; however, free-form accounts are difficult to analyze systematically in routine practice. This article reviews a special study, which automatically analyzed caregivers' open-ended responses to 12 questions inspired by the ADI-R in 51 families (children aged 2-8 years; 25 ASD and 26 controls). The study aimed to automatically detect ASD in children through computational analysis of parents' speech. The best subject-level classification strategy was achieved by converting text into semantic representations (using OpenAI's *text-ada-embedding-large-v3*), training models per question, including the question in the input, and aggregating decisions via majority voting

(84% accuracy; ROC-AUC 1.0). The clinical goal of this pilot study was to demonstrate the potential to support screening, prioritization, and referral, and potentially reduce waiting times and professional burden, thereby improving timely access to evaluation and early intervention. We discuss limitations (small sample, selected population, possible influence of prior intervention) and ethical risks (sociolinguistic bias, privacy, stigmatization, false positives/negatives), and we propose steps toward responsible clinical translation.

Keywords: autism spectrum disorder, early detection, caregiver narrative, computational analysis.

El trastorno del espectro autista (TEA) es un cuadro del neurodesarrollo con heterogeneidad marcada en presentación clínica, necesidades de apoyo y trayectorias evolutivas. El proceso diagnóstico requiere integrar múltiples fuentes: observación directa, historia del desarrollo y funcionamiento en contextos naturales. En la práctica, instrumentos estandarizados como la *AutismDiagnosticObservation Schedule* (ADOS)¹ y entrevistas con cuidadores en el marco de la *AutismDiagnostic Interview-Revised* (ADI-R)² suelen contribuir a esa integración. En primera línea, herramientas de cribado como el *Modified Checklist for Autism in Toddlers* (M-CHAT)³ permiten ordenar sospechas y orientar derivaciones, aunque no sustituyen la evaluación diagnóstica. Más allá de los instrumentos, el cuello de botella suele ser asistencial: en muchos servicios la demanda supera la capacidad disponible de diagnóstico, con listas de espera que desincronizan la sospecha inicial del acceso a evaluación especializada e intervención. En este escenario, herramientas que ayuden a organizar la demanda —sin reemplazar el acto diagnóstico— pueden aportar valor: priorizar derivaciones de mayor probabilidad, orientar entrevistas iniciales y decidir seguimientos más eficientes. En este comentario se analiza un estudio piloto que explora si el análisis automatizado del lenguaje de los cuidadores puede aportar una señal útil para apoyar el cribado y la priorización de derivaciones, sin sustituir el diagnóstico clínico⁴.

La narrativa parental en la sospecha del TEA

El relato del cuidador aporta una ventaja clínica evidente: recoge conductas del niño en

entornos donde la consulta no llega (rutinas, juego espontáneo, transiciones, situaciones de estrés y, a veces, contextos escolares). En el TEA, donde lo relevante suele vivir en el “cómo” y el “cuándo” de la interacción, las narrativas pueden aportar ejemplos y matices que un formato cerrado, como en el M-CHAT³, no captura bien. Aquí aparece un dilema práctico. Las preguntas cerradas son rápidas de administrar y puntuar, y facilitan comparabilidad entre sujetos. Las preguntas abiertas, en cambio, conceden autonomía al informante, reducen el sesgo de “respuesta sugerida” y capturan ejemplos espontáneos y heterogeneidad^{5,6}. El problema es que ese valor clínico se paga caro: transcripción, lectura cualitativa y síntesis consumen tiempo profesional, precisamente lo que suele escasear cuando crece la demanda asistencial. Desde la perspectiva de salud digital, distintas líneas de investigación han explorado cómo técnicas computacionales podrían aumentar la capacidad de análisis cuando el volumen de información supera la revisión humana sistemática⁷ y los biomarcadores de la patología son difíciles de detectar a través de la observación⁸.

Además, la narrativa parental no es un dato neutro: está mediada por cultura, nivel educativo, experiencias previas con el sistema de salud y el propio estilo comunicativo del cuidador. Existe literatura que sugiere diferencias en el lenguaje de padres de personas autistas respecto de personas con desarrollo típico, lo que apoya la hipótesis de que el discurso parental puede contener información discriminativa⁹. En salud, existen antecedentes de minería de texto aplicada a cuestionarios abiertos para reducir la carga de análisis, lo que sugiere viabilidad metodológica, aunque cada contexto clínico exige validación propia¹⁰. Hasta donde alcanza nuestro conocimiento, la literatura sobre el análisis computacional de los relatos de familiares de personas con TEA para mejorar los procesos de diagnóstico sigue siendo incipiente, con un único antecedente adicional (i.e., Mukherjee et al.¹¹) al estudio de Altozano et al.⁴.

El estudio de Altozano et al. (2025)

Altozano et al.⁴ evaluaron si el análisis automatizado de respuestas abiertas de progenitores a 12 preguntas específicas podía apoyar la cla-

sificación del TEA vs desarrollo típico. Participaron en el estudio 51 cuidadores de niños de 2 a 8 años (26 con hijos con desarrollo típico y 25 con TEA). En el grupo con TEA se incluyeron niños con diagnóstico confirmado y sin comorbilidades asociadas; además, los niños ya estaban en intervención. Por otra parte, los niños del grupo control no mostraban señales de desarrollo atípico.

Los padres contestaron por micrófono a 12 preguntas abiertas inspiradas en los dominios del ADI-R: interacción social, conductas repetitivas y lenguaje¹². Las preguntas se muestran en la **Tabla 1**.

Las respuestas a las preguntas se recogieron de forma oral para favorecer espontaneidad, en una sala tranquila, y luego se transcribieron incorporando rasgos no semánticos (p. ej., muletillas, risas, repeticiones). La **Figura 1** representa el *setup* experimental del estudio.

Debido a faltantes en algunas preguntas, se analizaron 593 respuestas individuales. Para minimizar el riesgo de detección del perfil de los ni

ños a partir de narrativas evidentes, los autores revisaron las transcripciones en busca de menciones explícitas de diagnóstico o participación en intervenciones. Aunque no se encontraron menciones directas, se registraron 11 referencias a “terapia” y “pictogramas”, términos plausiblemente asociados a intervención en TEA.

Desde una mirada clínica, el diseño tiene un punto fuerte: intenta formalizar algo que ocurre en la entrevista cotidiana (escuchar relatos y buscar consistencia entre dominios) en un formato breve y potencialmente escalable. Su aporte principal no es proponer un reemplazo del diagnóstico oficial, sino mostrar en condiciones controladas que la suma de narrativas breves estructuradas por dominios puede contener señal suficiente para apoyar procesos automáticos de cribado y priorización.

Análisis de los modelos computacionales

En el estudio⁴ se comparan varias formas de usar las narrativas abiertas de los padres para decidir si el perfil del niño se acerca más a TEA

Tabla 1 | Preguntas contestadas por los padres en el estudio de Altozano et al.⁴

Nº	Dominio	Pregunta
Q1	Interacción social	Piense en las expresiones no verbales de su hijo/a, como la mirada, la sonrisa, las expresiones faciales... ¿Cómo se comporta su hijo/a cuando está haciendo cosas con usted o cuando usted le habla?
Q2	Interacción social	¿Cómo interactúa y juega su hijo/a con otros niños, tanto conocidos como desconocidos?
Q3	Interacción social	¿Su hijo/a comparte cosas, juegos o juguetes con otros niños o con usted para jugar juntos? ¿Qué comparte?
Q4	Interacción social	¿Cómo muestra su hijo/a el deseo de algo y la necesidad de ayuda?
Q5	Interacción social	¿Cómo consigue establecer contacto con su hijo/a durante momentos de juego y en situaciones de estrés/necesidad?
Q6	Lenguaje	¿Cómo prefiere comunicarse su hijo/a con otros niños y con usted?
Q7	Interacción social	¿Qué juegos imitativos y/o imaginativos juega su hijo/a?
Q8	Lenguaje	¿Cómo se comporta su hijo/a en interacciones verbales con otros niños y con usted?
Q9	Lenguaje	¿Su hijo/a utiliza frases o hace sonidos de forma repetitiva? ¿Cuáles son esas frases?
Q10	Conductas repetitivas	¿Su hijo/a tiene intereses en objetos, cosas o juegos que repite constantemente? Si es así, ¿cuáles?
Q11	Conductas repetitivas	¿Su hijo/a tiene rituales o secuencias fijas de acciones que tiene que cumplir? Si es así, ¿cuáles?
Q12	Conductas repetitivas	¿Su hijo/a muestra movimientos repetitivos o acciones estereotipadas que aparecen a lo largo del día y en diferentes situaciones? Si es así, ¿cuáles?

Figura 1 | Setup experimental del estudio de Altozano et al.⁴

o a desarrollo típico. Por un lado, los autores prueban dos enfoques clásicos de aprendizaje automático: uno que convierte cada texto en una representación numérica ya preparada (un modelo de “embeddings” de OpenAI) y otro basado en un modelo de lenguaje tipo RoBERTa. En ambos casos, el sistema aprende un clasificador final para separar TEA vs desarrollo típico y se evalúa en dos variantes sencillas de entrada: usar solo la respuesta del padre/madre, o incluir también la pregunta junto con la respuesta para dar más contexto.

Además, con RoBERTa comparan dos maneras de “ajustar” el modelo: una más conservadora (*frozen*, se deja el modelo base tal cual y se entrena solo la capa final) y otra más adaptable (*unfrozen*, se permite que el modelo se ajuste en conjunto a este problema). También los autores evalúan distintas estrategias prácticas: entrenar un único modelo con todas las preguntas del cuestionario, entrenar modelos agrupando las preguntas por dominios, o entrenar un modelo por pregunta; y para obtener el resultado final por niño, combinan la información de las 12 respuestas promediando las probabilidades que produce el sistema. Por último, incluyen una comparación con GPT-4o, sin entrenamiento (*zero-shot*), al que se le presentan las pregun-

tas y respuestas como un intercambio estructurado y se le pide que valore cada respuesta en una escala de 0 a 6 según indicios de TEA y, con el conjunto completo, emita una decisión final entre las dos clases, evaluándolo en los mismos conjuntos de prueba que el resto.

El mejor rendimiento a nivel de sujeto se obtuvo con una estrategia que transforma cada respuesta en una representación numérica del significado del texto (“*embeddings*”, *text-ada-embedding-large-v3*), entrena un clasificador específico por cada pregunta, incorpora explícitamente el texto de la pregunta junto con la respuesta y combina las 12 predicciones mediante una agregación tipo “votación”. Con esta configuración se reportó 84% de precisión, ROC-AUC 1.0, sensibilidad 72% y especificidad 96%.

Los autores proponen que el entrenamiento por pregunta puede actuar como una forma de regularización en un escenario con pocos sujetos: cada modelo “especialista” aprende dentro de un alcance estrecho (una pregunta/un dominio) y la agregación posterior mejora rendimiento y estabilidad. También encuentran que incluir la pregunta explícitamente tiende a mejorar o mantener el desempeño, sobre todo cuando las respuestas son breves o dependientes del contexto (p. ej., “no”, “a veces”). En términos clínicos,

esto puede entenderse como “forzar contexto”: una respuesta telegráfica solo adquiere significado si se sabe qué conducta estaba siendo explorada. Además, la combinación por votación podría amortiguar el peso de un ejemplo aislado y favorecer patrones consistentes entre áreas (p. ej., comunicación social y juego). Estas hipótesis requieren validación externa y pruebas en escenarios reales.

Implicaciones clínicas del estudio

Aunque los resultados del estudio de Altozano et al.⁴ son muy prometedores, siguen siendo preliminares y requieren una validación más robusta. Si se valida el modelo en muestras más grandes y diversas, este enfoque podría integrarse como apoyo en contextos de alta demanda: tras una breve entrevista abierta (presencial o remota), generar un indicador de riesgo que ayude a priorizar derivaciones y organizar circuitos. Su atractivo es pragmático: convierte en señal cuantificable una materia prima ya disponible en la clínica (narrativa parental) y puede favorecer acceso cuando la logística o la geografía dificultan entrevistas especializadas. En sistemas con recursos limitados, un uso prudente podría contribuir a acortar la distancia entre sospecha inicial y evaluación, mejorando la coordinación de circuitos asistenciales. Sin embargo, esto no habilita a “diagnosticar por texto”. Sí sugiere que, cuando la narrativa parental se recoge de modo abierto pero estructurado por dominios, puede contener información útil para apoyar decisiones de cribado, priorización de derivación o seguimiento intensivo. El matiz es central: su lugar clínico razonable, si se valida, es ordenar demanda y orientar el acceso a evaluación, no reemplazar instrumentos ni juicio clínico.

Riesgos, límites y consideraciones éticas

Un primer riesgo inevitable en cualquier herramienta de apoyo al cribado es el equilibrio entre falsos positivos y falsos negativos. Un falso positivo puede incrementar la ansiedad familiar, favorecer la estigmatización y generar derivaciones innecesarias; en cambio, un falso negativo puede retrasar la evaluación clínica y, con ello, el acceso a intervención temprana. En el mejor modelo del estudio, se observó alta especificidad y sensibilidad moderada, un patrón que obliga

a pensar no solo en la calibración de umbrales, sino en la comunicación del resultado: conviene hablar de “riesgo” y “recomendación de evaluación” en lugar de etiquetas categóricas. En paralelo, el principal límite es la validez externa. La generalización queda restringida por una muestra pequeña, criterios de inclusión restrictivos y la ausencia de población no diagnosticada, lo que impide extrapolar el rendimiento a escenarios reales de diagnóstico diferencial⁴. Además, el grupo TEA ya estaba en intervención, un factor que puede modular la narrativa parental y, por tanto, el desempeño del sistema. Esto refuerza un punto clínico central: la herramienta debe validarse exactamente allí donde pretende usarse, idealmente en niños aún no diagnosticados, antes de integrarse en decisiones asistenciales. La privacidad y la gobernanza del dato constituyen otra dimensión crítica, dado que se trabaja con información sensible de menores y familias. El propio planteamiento introduce un *trade-off* relevante: componentes propietarios de gran escala pueden rendir bien, pero a menudo implican procesamiento en la nube; en contraste, soluciones locales maximizan el control de los datos y facilitan la auditoría. Cualquier despliegue con aspiración clínica debería explicitar, de forma trazable y comprensible, qué datos se almacenan, durante cuánto tiempo, quién accede, con qué fines secundarios (si los hubiera) y cómo se ejerce el derecho a retirar el consentimiento. Por último, la equidad no viene incorporada de fábrica. Si el modelo aprende estilos narrativos asociados a determinados grupos (por nivel educativo, dialecto o bilingüismo), podría priorizar de manera desigual. Por ello, la validación debe incluir diversidad sociocultural y lingüística, y reportar desempeño por subgrupos relevantes antes de que el sistema influya, siquiera indirectamente, en decisiones asistenciales.

En términos de traslación clínica, antes de considerar cualquier uso en circuito asistencial se requieren etapas claras y secuenciales. En primer lugar, una validación prospectiva en población no diagnosticada, comparando el indicador con el cribado habitual y con el diagnóstico clínico final⁴. En segundo lugar, la inclusión de comorbilidades y condiciones frecuentes (p. ej., trastornos del lenguaje, TDAH, ansiedad, discapacidad intelectual), dado que la práctica rara

vez ofrece perfiles TEA “puros”⁴. En tercer lugar, la generalización cultural y lingüística, evaluando el desempeño en distintas variedades del español y contextos socioculturales, con análisis por subgrupos y revisión activa de sesgos. En cuarto lugar, la definición del “producto clínico”: no basta con devolver una probabilidad, sino que se necesitan salidas interpretables para el equipo (qué dominios contribuyeron más, ejemplos textuales relevantes) y mensajes comprensibles para las familias centrados en priorización de evaluación, no en diagnóstico. Finalmente, la integración debería ser supervisada y protocolizada en el lugar clínico más plausible, que es el *triage* (priorización de turnos/derivaciones) en admisión, atención primaria o teleconsulta, con capacitación del equipo, auditoría continua y resguardo de datos⁴. Si el objetivo final es aliviar carga asistencial y mejorar acceso, la evaluación de impacto debería ir más allá de métricas de clasificación e incluir resultados operativos y humanos: tiempos de espera, tasa de derivaciones, aceptabilidad familiar, seguridad y efectos no deseados (p. ej., ansiedad o estigma). En otras

palabras, una herramienta puede “rendir bien” en exactitud y, aun así, causar daño si no se integra con cuidado, supervisión clínica y una comunicación responsable del riesgo.

Conclusión

El análisis automatizado del lenguaje de los padres, recogido mediante preguntas abiertas estructuradas, podría convertirse en una señal escalable para apoyar cribado y priorización de derivaciones en TEA. El mensaje global es clínico: puede ayudar a identificar señales precoces y a ordenar circuitos de atención, sin sustituir el diagnóstico, que sigue siendo un acto profesional contextual. La prudencia es obligada: el estudio de Altozano et al.⁴ se trata de un estudio piloto con límites claros de generalización y con desafíos éticos y de gobernanza del dato. Si se avanza con validación rigurosa, transparencia y supervisión clínica, este tipo de herramientas podría contribuir a un circuito más ágil y equitativo, con menos fricción para familias y servicios.

Conflicto de intereses. Ninguno para declarar.

Bibliografía

1. Lord C, Rutter M, DiLavore PC, et al. *Autism Diagnostic Observation Schedule (ADOS). Manual*. Los Angeles: Western Psychological Services; 1999.
2. Lord C, Rutter M, Couteur L, et al. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 1994; 24: 659-85.
3. Robins D, Fein D, Barton M. *Modified Checklist for Autism in Toddlers (M-CHAT) Follow-Up Interview*. Publisher: Author; 1999.
4. Altozano A, Minissi ME, Gomez-Zaragoza L, et al. Enhancing Psychological Assessments with Open-Ended Questionnaires and Large Language Models: An ASD Case Study. *IEEE J Biomed Health Inform* 2025 Aug 15;PP. doi: 10.1109/JBHI.2025.3599643.
5. Foddy W. *The open vs. closed questions debate*. Cambridge: Cambridge University Press; 1993:126-52.
6. Reja U, Manfreda K, Hlebec V, et al. Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics* 2003; 19:159-77.
7. Wang X-J, Krystal JH. Computational psychiatry. *Neuron* 2014; 84:638-54.
8. Minissi ME, Altozano A, Marín-Morales J, et al. Biosignal comparison for autism assessment using machine learning models and virtual reality. *Computers in Biology and Medicine* 2024; 171:108-94.
9. Rahman MM, Usman OL, Muniyandi RC, et al. A review of machine learning methods of feature selection and classification for autism spectrum disorder. *Brain Sci* 2020;10(12):949. doi: 10.3390/brainsci10120949.
10. Landa R, Piven J, Wzorek MM, et al. Social language use in parents of autistic individuals. *Psychological Medicine* 1992; 22:245-54.
11. Mukherjee P, Sadhukhan S, Godse M, et al. Early Detection of Autism Spectrum Disorder (ASD) using Traditional Machine Learning Models. *International Journal of Advanced Computer Science and Applications* 2023; 14: 231-45.
12. Giglioli IAC, Maddalon L, Gómez-Zaragoza L, et al. A voice recognition application for the semantic and prosodic analysis of ASD caregivers. *Annual Review Of Cybertherapy And Telemedicine* 2021; 53.