



GOBIERNO
DE ESPAÑA

MINISTERIO
DE DERECHOS SOCIALES, CONSUMO
Y AGENDA 2030

SECRETARÍA DE ESTADO
DE DERECHOS SOCIALES



CREAP

Centro de
Referencia
Estatad de
Atención
Psicosocial

INFORME DE TENDENCIAS TRASTORNO MENTAL GRAVE

Diciembre 2025

Informe mensual a 31 de diciembre de 2025

Índice

ÍNDICE	1
TENDENCIAS CIENTÍFICAS EN SALUD MENTAL Y TMG	2
Manuscritos en Research Square	2
Manuscritos en OSF Preprints	10
Manuscritos en MedRxiv y BioRxiv	20
Manuscritos en Preprints.org	32

Tendencias científicas en salud mental y TMG

NOTA: Estos artículos son preimpresiones, es decir, manuscritos originales que no han sido revisados por pares. Informan de nuevas investigaciones que aún no han sido evaluadas por una serie de estándares de calidad técnica y científica.

Periodo de búsqueda: 1 de noviembre al 31 de diciembre de 2025.

Búsqueda principal: severe mental disorder OR severe mental illness.

Búsquedas secundarias: schizophrenia, psychosis, mental health, bipolar disorder, depression, borderline personality disorder.

Manuscritos en Research Square

Plataforma de pre-prints multidisciplinaria que forma parte de Springer Nature desde diciembre de 2022. <https://www.researchsquare.com/>

Evaluating a brief MBCT programme for non-suicidal self-injury in individuals with BPD: A quasi-experimental, nonrandomised pre-post design pilot study

Szilvia Kresznerits, Ágnes Zinner-Gérecz, Mónika Miklósi, Tamás Szekeres

Borderline personality disorder, non-suicidal self-injury, mindfulness-based cognitive therapy, emotion regulation, feasibility study

Introducción: La autolesión no suicida es muy frecuente en personas con trastorno límite de la personalidad y constituye uno de los predictores más sólidos de futuros intentos de suicidio y de muerte por suicidio. El acceso a tratamientos integrales como la terapia dialéctico-conductual sigue siendo limitado en muchos contextos clínicos.

Método: Este estudio exploratorio evaluó la viabilidad y los resultados preliminares de un programa breve de terapia cognitiva basada en mindfulness adaptado para reducir la autolesión no suicida y mejorar procesos psicológicos asociados. Se utilizó un diseño no aleatorizado pre-post con comparador intragrupo mediante lista de espera. Las evaluaciones se realizaron en tres momentos: entre 8 y 12 semanas antes de la intervención (T1; n = 120), inmediatamente antes de la intervención (T2; n = 72) y tras la intervención (T3; n = 50). Se reclutó a pacientes ambulatorios con diagnóstico de trastorno límite de la personalidad y autolesión reciente para participar en un programa grupal de 9 semanas. Los criterios de exclusión incluyeron psicosis activa, episodios maníacos, consumo grave de sustancias y riesgo suicida agudo. Las variables principales fueron la frecuencia de autolesión no suicida y el mindfulness. Las variables secundarias incluyeron autocompasión, autoestima, impulsividad,



depresión, disociación y regulación emocional. Los cambios intragrupo se analizaron mediante modelos lineales mixtos y pruebas de rangos con signo de Wilcoxon. Se aplicó regresión logística binaria para identificar predictores de abandono. No se utilizaron procedimientos de cegamiento ni aleatorización.

Resultados: Las personas que completaron la intervención mostraron reducciones significativas en la frecuencia de autolesión, los síntomas depresivos, la desesperanza, la impulsividad y la regulación emocional desadaptativa, junto con mejoras en mindfulness, autocompasión y autoestima. Estos cambios se observaron principalmente durante la fase de intervención y no durante el periodo en lista de espera, lo que sugiere efectos específicos del tratamiento. Los síntomas disociativos mostraron una tendencia a la mejora que no alcanzó significación estadística. El abandono entre los pacientes elegibles se asoció con mayor impulsividad, regulación emocional desadaptativa y, de forma inesperada, mayor autocompasión, mientras que un menor número de diagnósticos comórbidos predijo una menor retención en el tratamiento.

Conclusión: El programa MBCT adaptado a la autolesión no suicida podría ser una intervención complementaria viable y potencialmente eficaz para personas con trastorno límite de la personalidad, al mejorar la regulación emocional y reducir la frecuencia de autolesión. Los resultados respaldan la necesidad de futuros estudios controlados y con seguimientos más prolongados.

Background: Non-suicidal self-injury (NSSI) is highly prevalent among individuals with borderline personality disorder (BPD) and is one of the most robust predictors of future suicide attempts and suicide deaths. Access to comprehensive treatments such as dialectical behaviour therapy remains limited.

Methods: This exploratory study evaluated the feasibility and preliminary outcomes of a brief, group-based mindfulness-based cognitive therapy program tailored to reduce NSSI (MBCT-NSSI) and improve associated psychological processes. Using a nonrandomised pre–post design with a within-subject waitlist control comparator, assessments were conducted at three time points: 8–12 weeks pre-intervention (T1) (n = 120), immediately pre-intervention (T2) (n = 72), and post-intervention (T3) (n = 50). Outpatients diagnosed with BPD and recent NSSI were recruited from psychiatric clinics to participate in a 9-week MBCT-NSSI group programme. Exclusion criteria included current psychosis, manic episodes, severe substance use, and acute suicide risk. Primary outcomes were NSSI frequency and mindfulness. Secondary outcomes included self-compassion, self-esteem, impulsivity, depression, dissociation, and emotion regulation. Within-subject changes were analysed via linear mixed modelling and Wilcoxon signed-rank tests. Binary logistic regression identified predictors of dropout among eligible participants. No blinding or randomisation procedures were applied.



Results: Participants who completed the MBCT-NSSI intervention (N = 50) presented significant reductions in NSSI frequency, depressive symptoms, hopelessness, impulsivity, and maladaptive emotion regulation, alongside improvements in mindfulness, self-compassion, and self-esteem. These changes were observed primarily during the intervention phase and not during the waiting-list period, suggesting treatment-specific effects. Dissociative symptoms showed a non-significant trend toward improvement. Dropout among eligible patients (N = 120) was associated with higher impulsivity, maladaptive regulation, and, unexpectedly, higher self-compassion, whereas fewer comorbid diagnoses predicted lower treatment retention.

Conclusions: MBCT-NSSI may be a feasible and effective adjunctive intervention for individuals with BPD who engage in NSSI, improving emotion regulation and reducing self-harm frequency. These findings support further investigations via controlled trials and longer follow-up periods.

Publicado: 1 de diciembre de 2025

Cytokine levels in symptomatic patients undergoing anti-psychotic treatment for First Episode Psychosis: a cross-sectional study.

Valentina Morgera, Frederick Sundram, Bruce Russell, Ian Soosay

BMC Psychiatry

Introducción: En la psicosis de primer episodio, las citocinas parecen activarse como parte de un Sistema de Respuesta Inflamatoria (IRS) o de un Sistema de Respuesta Inflamatoria Compensatoria (CIRS) y se han propuesto como marcadores clínicos, tanto de estado como de rasgo, de la psicosis. El tratamiento antipsicótico tiende a normalizar el equilibrio IRS/CIRS, aunque persiste una activación predominante del IRS. Este estudio analizó los niveles sanguíneos de citocinas IRS y CIRS propuestas como marcadores clínicos en pacientes con psicosis de primer episodio que presentaban síntomas y recibían tratamiento antipsicótico, controlando de forma exhaustiva factores y comorbilidades asociados a cambios inmunológicos.

Método: Se compararon los niveles séricos de 16 citocinas en pacientes con psicosis de primer episodio (n = 20), reclutados en un servicio clínico especializado, con un grupo control sano emparejado por edad, sexo e índice de masa corporal (n = 20). Se controló el consumo de tabaco y cannabis, la medicación concomitante y la ausencia de fiebre. Se aplicaron modelos de regresión con mejor ajuste para identificar diferencias entre pacientes y controles, considerando además edad, sexo e IMC. También se analizaron las correlaciones entre los niveles de citocinas, las puntuaciones en la Positive and Negative Syndrome Scale (PANSS) y la carga antipsicótica, expresada en equivalentes de clorpromazina.



Resultados: En comparación con los controles sanos, los pacientes con psicosis de primer episodio mostraron niveles elevados de citocinas del IRS (IL-2 e IL-6) y niveles reducidos de otras citocinas del IRS (IL-1, IL-8, IL-12) y del CIRS (IL-4). No se observaron diferencias significativas en la ratio IRS/CIRS ni correlaciones entre los niveles de citocinas y las puntuaciones de la PANSS. Se encontraron correlaciones negativas entre niveles más bajos de IL-2, IL-6 e IL-8 y mayores dosis de antipsicóticos.

Conclusión: En pacientes con psicosis de primer episodio, clínicamente estables y bajo tratamiento antipsicótico, la ratio IRS/CIRS no se mostró alterada, aunque sí se detectaron cambios específicos en algunas citocinas. IL-2 e IL-6 permanecieron elevadas pese al efecto inmunomodulador de los antipsicóticos. Estos hallazgos sugieren que podrían utilizarse, junto con la proteína C reactiva, como posibles marcadores de estado y/o rasgo para identificar a personas con psicosis de primer episodio que podrían beneficiarse de terapias antiinflamatorias adyuvantes.

Background: Cytokines in first-episode psychosis (FEP) appear to be activated as part of an Inflammatory Response System (IRS) or a Compensatory Inflammatory Response System (CIRS) and have been proposed as clinical (state and trait) markers of psychosis. Antipsychotic medication normalizes the IRS/CIRS ratio despite a predominant activation of the IRS. The current study examined blood levels of IRS/CIRS cytokines proposed as clinical markers in symptomatic medicated patients with FEP. We screened all co-diagnoses and immunomodulating factors that have been linked to cytokine changes to increase the accuracy of findings.

Methods: We compared cytokines (n = 16) levels in the blood serum of patients with FEP (n = 20) recruited from a specialized FEP clinical service with a matched (age, gender, body mass index) group of healthy control subjects (HC) (n = 20). We screened for tobacco smoking, cannabis use, medications, and the absence of fever. Best-fitting regression models were used to identify differences in FEP vs HCs per diagnosis and additional variables of age, gender, and BMI. We correlated cytokine levels to Positive and Negative Symptoms Scale (PANSS) scores and antipsychotic load, recorded as chlorpromazine equivalents.

Results: We found elevated levels of IRS (IL-2 and IL-6), and reduced levels of IRS (IL-1, IL-8, IL-12), and CIRS (IL-4) in FEP vs HC. No significant differences in the IRS/CIRS ratio or a correlation between cytokine levels and PANSS scores were found. Lower levels of IL-2, IL-6 and IL-8 correlated to higher amounts of antipsychotics.

Conclusion: Our study on well-controlled moderately symptomatic patients with FEP taking antipsychotic treatment found that the IRS/CIRS ratio was not altered, while some alterations in cytokine levels were found. IL-2 and IL-6 remain elevated despite the immune-modulatory action of the antipsychotic agents; we suggest they could be used as screening tools for either trait (IL-6) and/or state (IL-2) markers of psychosis, in combination with C-reactive protein, to identify people with FEP that



could potentially benefit from an adjuvant anti-inflammatory therapy alongside antipsychotics.

Publicado: 24 de noviembre de 2025

Understanding Relationships Between Birth Trauma, Helplessness, and Perinatal Mental Health Disorders

Stephanie Hanson

Perinatal mental health, birth trauma, helplessness, postpartum women, psychological mechanisms

Introducción: El trauma obstétrico y los trastornos de salud mental perinatal constituyen un problema relevante de salud pública. Sin embargo, las relaciones entre ambos y los mecanismos psicosociales que los conectan siguen poco explorados. Este estudio analizó la asociación entre el trauma durante el parto, los sentimientos de indefensión y los trastornos de salud mental perinatal en mujeres en el posparto.

Método: Se realizó una encuesta transversal a mujeres residentes en Estados Unidos entre los 4 y 7 meses tras el parto, reclutadas a través de Amazon Mechanical Turk. Las participantes elegibles (N = 249) completaron instrumentos validados para evaluar trauma obstétrico, indefensión y distintos trastornos de salud mental perinatal, incluyendo depresión, ansiedad, trastorno obsesivo-compulsivo, trastorno bipolar y psicosis. Se aplicaron modelos de regresión de Poisson modificada y modelos lineales, ajustados por variables sociodemográficas y obstétricas.

Resultados: El 21,3% de las participantes informó haber experimentado trauma obstétrico y el 21,7% refirió sentimientos de indefensión. La prevalencia de trastornos de salud mental perinatal fue elevada: trastorno obsesivo-compulsivo (93,6%), depresión (84,7%), ansiedad (81,9%), trastorno bipolar (10%) y psicosis (3,2%). Todas las mujeres que reportaron trauma obstétrico presentaron al menos un trastorno de salud mental perinatal en el cribado. El trauma obstétrico se asoció de forma positiva con la indefensión (RR ajustado = 1,40; IC 95%: 1,18–1,66). Mayores niveles de indefensión predijeron un aumento del riesgo y de la gravedad de los síntomas de depresión (RR = 1,10), ansiedad (RR = 1,13) y trastorno obsesivo-compulsivo (RR = 1,05). Los modelos lineales confirmaron asociaciones con todos los trastornos analizados. Las asociaciones con trastorno bipolar y psicosis fueron más débiles e inconsistentes.

Conclusión: Los resultados señalan la indefensión como un posible mecanismo psicológico que conecta el trauma obstétrico con diversos trastornos de salud mental perinatal. La incorporación de una atención perinatal informada en trauma y orientada al empoderamiento, junto con el cribado sistemático de trauma obstétrico e indefensión, podría reducir la psicopatología en el posparto. Estudios longitudinales



futuros ayudarán a clarificar las relaciones causales y a desarrollar intervenciones específicas.

Background: Birth trauma and perinatal mental health disorders (PMHDs) are significant public health concerns, yet their interrelationships and underlying psychosocial mechanisms remain underexplored. This study examined associations between birth trauma, feelings of helplessness, and PMHDs among postpartum women.

Methods: A cross-sectional survey was administered to U.S. women 4–7 months postpartum via Amazon Mechanical Turk. Eligible participants (N = 249) completed validated measures assessing birth trauma, helplessness, and PMHDs, including depression, anxiety, obsessive-compulsive disorder (OCD), bipolar disorder, and psychosis. Modified Poisson and linear regression models evaluated associations, adjusting for sociodemographic and obstetric factors.

Results: Birth trauma was reported by 21.3% of participants, and 21.7% reported feelings of helplessness. PMHD prevalence was high: OCD (93.6%), depression (84.7%), anxiety (81.9%), bipolar disorder (10%), and psychosis (3.2%). All participants reporting birth trauma screened positive for at least one PMHD. Birth trauma was positively associated with helplessness (adjusted RR = 1.40; 95% CI: 1.18–1.66). Higher helplessness scores predicted increased risk and symptom severity for depression (RR = 1.10), anxiety (RR = 1.13), and OCD (RR = 1.05), with linear models confirming associations across all PMHDs. Associations with bipolar disorder and psychosis were weaker and inconsistent across models.

Conclusions: Findings highlight helplessness as a potential psychological mechanism linking birth trauma to diverse PMHDs. Integrating trauma-informed, empowerment-oriented care into perinatal practice and routine screening for birth trauma and helplessness may mitigate postpartum psychopathology. Future research should employ longitudinal designs to clarify causal pathways and inform targeted interventions.

Publicado: 23 de noviembre de 2025

Do not stop me now: A feasibility study for a personalized, phased psychomotor sports intervention to promote social integration in people with early psychosis.

Merel Karlijn Muller, Akke-Marij D. Ariesen, Jooske T. Busschbach

Personalized sports intervention, psychomotor therapy, social integration, early psychosis, feasibility study

Introducción: La integración social suele verse comprometida en personas con psicosis temprana. Las intervenciones basadas en la actividad física y el deporte pueden favorecer la participación social, pero existe poca evidencia sobre programas personalizados y progresivos adaptados a esta población. Este estudio evaluó la



viabilidad de una intervención deportiva psicomotriz personalizada y por fases como paso previo a un estudio de mayor tamaño.

Método: Tres participantes realizaron una intervención de 20 semanas orientada a promover la (re)integración social a través del deporte. El programa incluyó tres fases: terapia psicomotriz individual, terapia psicomotriz grupal (ambas en un entorno de salud mental) y participación deportiva en la comunidad. Se empleó un diseño de línea base múltiple de caso único, junto con entrevistas cualitativas, el Experience Sampling Method (ESM), acelerómetros ActiGraph GT9X+ y cuestionarios estandarizados.

Resultados: El proceso de reclutamiento resultó complejo, ya que solo se incluyeron tres personas de veinte candidatas potencialmente elegibles. Las medidas de resultado seleccionadas fueron en general viables, aunque se requirieron ajustes, especialmente en el procedimiento ESM. La intervención fue valorada positivamente en términos de aceptabilidad, aplicabilidad y beneficios individuales percibidos. Se formularon recomendaciones sobre la duración del programa y sobre el papel del profesional de apoyo entre iguales y del entrenador. Al finalizar el estudio, todas las personas participantes realizaban algún tipo de actividad física, como yoga, boxeo o caminatas regulares.

Conclusión: Con criterios de inclusión más amplios y mayor flexibilidad en la recogida de datos para adaptarse a las necesidades individuales, esta investigación de viabilidad respalda el potencial de una intervención deportiva personalizada y por fases para mejorar la integración social en personas con psicosis temprana. Los resultados aportan orientaciones metodológicas clave para futuros estudios de caso único con diseño de línea base múltiple a mayor escala.

Purpose: This study examines the feasibility of a personalised and phased psychomotor sports intervention in preparation for a larger multiple-baseline single-case design study aimed at enhancing social integration in adults with early psychosis. The aim is to evaluate the recruitment strategy, the suitability of selected outcome measures, and the applicability of the intervention for the target population.

Methods: Three participants engaged in a 20-week intervention designed to promote social (re-) integration through sports participation, with three phases 1) individual psychomotor therapy, (2) group psychomotor therapy, both offered within a mental health care setting and (3) community-based sports participation. Next to (qualitative) interviews, a multiple-baseline design was used, incorporating the Experience Sampling Method (ESM), ActiGraph GT9X+ accelerometers, and questionnaires.

Results: Recruitment proved challenging, with only three participants included out of twenty potentially suitable candidates. The selected outcome measures were largely feasible, with some adjustments made, particularly in the ESM procedure. The intervention was evaluated positively in terms of acceptability, applicability, and



potential individual benefits, with several recommendations made regarding its duration and the roles of the peer support worker and the coach. All participants engaged in some kind of physical activity by the end of the study (yoga, boxing or regular walking).

Conclusion: With broader inclusion criteria and flexibility in the data gathering procedure to tailor individual needs, this feasibility study supports the potential of a personalised, phased sports intervention to enhance social integration in early psychosis and informs methodological refinements for a future, larger single-case multiple-baseline study.

Publicado: 6 de noviembre de 2025



Manuscritos en OSF Preprints

Plataforma multidisciplinaria de acceso abierto gestionada por el Center for Open Science. Actúa como agregador de servidores temáticos (p. ej., PsyArXiv, SocArXiv) y permite también subir manuscritos al repositorio general.
https://osf.io/preprints?view_only

Understanding belief formation and delusion-like phenomena in human–AI interaction: a belief-updating perspective

Hamilton Morrin, Quinton Deeley, Thomas Pollak

Artificial intelligence, belief updating, delusions, epistemic drift, human–AI interaction

Introducción: Los sistemas contemporáneos de inteligencia artificial generativa, cada vez más ajustados a la cognición social humana, están pasando a desempeñar un papel activo en la forma en que las personas forman y revisan sus creencias. En el ámbito clínico han empezado a describirse presentaciones centradas en delirios asociados al uso de IA, cuyo inicio o contenido parece vinculado a diálogos prolongados con modelos de lenguaje. Al mismo tiempo, muchas personas usuarias refieren cambios subclínicos en sus convicciones y experiencias de tipo “revelatorio” que modifican su conducta y su visión del mundo sin cumplir criterios de psicosis.

Marco teórico: Los autores proponen que estos fenómenos pueden entenderse desde un marco de actualización de creencias, en el que las respuestas de los agentes de IA funcionan como testimonio sobre el mundo. Las configuraciones de interacción —que influyen, por ejemplo, en la memoria o en la postura interpersonal— modulan el peso o la precisión asignada a ese testimonio.

Desarrollo conceptual: Desde esta perspectiva, los delirios asociados a la IA se sitúan en el extremo de un continuo más amplio de efectos epistémicos, que va desde un desplazamiento gradual de creencias hasta convicciones altamente cristalizadas. Este continuo se inserta en una ecología de creencias más amplia, configurada por diadas personalizadas entre humanos y sistemas de IA. Los autores desarrollan una analogía de “psicofarmacología virtual”, en la que distintas configuraciones de sistemas de IA producen efectos sobre la dinámica de las creencias comparables a cambios neuromoduladores en la precisión asignada a la evidencia social. Asimismo, se analiza cómo los estados dopaminérgicos y otros estados neurobiológicos podrían modificar la susceptibilidad al diálogo que desplaza creencias.

Implicaciones clínicas y sociales: El trabajo examina cómo agentes deliberadamente configurados podrían utilizarse para apoyar el bienestar en contextos clínicos definidos. Al mismo tiempo, analiza cómo esas mismas configuraciones pueden emplearse para moldear creencias y atención a gran escala, por ejemplo en



productos diseñados para inducir estados espirituales o epifánicos, como mecanismos de radicalización en contextos extremistas o sectarios, o en clones de personalidad con fines políticos.

Conclusión: Los autores concluyen que las configuraciones de interacción deben considerarse influencias modificables sobre la creencia y la atención. La gobernanza de estos sistemas debería abordar tanto la concentración del control sobre estos “reguladores” como los sesgos estructurales que determinan qué perspectivas testimoniales se amplifican y cuáles se suavizan o se ignoran.

Background: Contemporary generative AI systems, increasingly adapted to human social cognition, are becoming active participants in how people form and revise beliefs. Clinicians have begun to describe AI-associated delusion-centred presentations whose onset or content appears closely associated with users’ extended dialogue with large language models. Many more users report subclinical shifts in conviction and “revelatory” experiences that alter behaviour and world-view without meeting criteria for psychosis.

Framework: We argue that these phenomena can be understood within a belief-updating framework in which AI agents’ outputs function as testimony about the world, and their interaction configurations (that shape, for example, memory and interpersonal stance) weight the precision of that testimony.

Conceptual development: On this view AI-associated delusions sit at the extreme of a broader spectrum of epistemic effects, extending from gradual epistemic drift to highly crystallised conviction, situated within a wider ecology of belief shaped by personalised human-AI dyads. We develop a virtual psychopharmacology analogy in which different AI system configurations have effects on belief dynamics that resemble neuromodulatory changes in the precision assigned to social evidence, and we consider how dopaminergic and related states may alter susceptibility to belief-shifting dialogue.

Implications: We consider how deliberately configured agents might be used to support wellbeing in defined clinical contexts and analyse how the same configurations could be deployed to shape belief and attention at population scale, including in products designed to induce spiritual or epiphanic states, as mechanisms of radicalisation in extremist or cultic contexts, and in persona clones used for political purposes.

Conclusion: We conclude that interaction configurations should be treated as modifiable influences on belief and attention, and that governance must address both the concentration of control over these dials, as well as the structural biases that determine whose testimonial perspectives are amplified and whose are smoothed over or ignored.

Publicado: 19 de diciembre de 2025



From Voice to Self: An Integrative Framework on Self-Voice Processing

Pavo Orepic, Ana Pinheiro

Self-voice processing, self-perception, multisensory integration, psychosis, voice technology

Introducción: La voz propia desempeña un papel central en la comunicación y en la identidad personal, pero ha recibido relativamente poca atención en la psicología científica. La expansión de voces generadas por inteligencia artificial y de voces manipuladas digitalmente hace cada vez más relevante comprender cómo las personas perciben y procesan su propia voz. Además, las alteraciones en el procesamiento de la voz propia están implicadas en diversas condiciones clínicas, como la psicosis, el autismo y los trastornos de la personalidad, lo que subraya la necesidad de modelos integradores.

Retos conceptuales y metodológicos: La investigación en este ámbito se enfrenta a dos dificultades principales. Por un lado, un reto metodológico relacionado con la reproducción de la acústica por conducción ósea que caracteriza la percepción natural de la voz propia. Por otro, un reto conceptual derivado de la tendencia persistente a tratar la voz propia como un fenómeno exclusivamente auditivo, dejando de lado otros procesos implicados.

Marco teórico propuesto: Para abordar estas limitaciones, las autoras proponen un marco integrador que descompone la voz propia en cinco componentes interrelacionados: auditivo, motor, memoria, integración multisensorial y autoconcepto. Cada uno de estos componentes contribuye de manera específica al procesamiento coherente de la voz propia.

Bases funcionales y neurales: El trabajo revisa la base funcional y neural de cada componente y plantea cómo convergen dentro de redes cerebrales distribuidas. Esta convergencia permitiría sostener una experiencia unificada y estable de la voz propia en distintos contextos.

Conclusión: El marco integrador propuesto pretende impulsar el desarrollo teórico y traslacional al conectar la psicología, la neurociencia, la investigación clínica y la tecnología de la voz. De este modo, ofrece una base conceptual útil para comprender la voz propia en entornos digitales emergentes y para futuras aplicaciones clínicas y tecnológicas.

Background: The self-voice plays a fundamental role in communication and identity, yet remains a relatively neglected topic in psychological science. As AI-generated and digitally manipulated voices become more common, understanding how individuals perceive and process their own voice is increasingly important. Disruptions in self-voice processing are implicated in several clinical conditions, including psychosis, autism, and personality disorders, highlighting the need for integrative models to



explain self-voice across contexts.

Challenges: However, research faces two major challenges: a methodological one – replicating the bone-conducted acoustics that shape natural self-voice perception, and a conceptual one – a persistent bias toward treating the self-voice as purely auditory.

Framework: To address these gaps, we propose a framework decomposing the self-voice into five interacting components: auditory, motor, memory, multisensory integration, and self-concept.

Neural basis: We review the functional and neural basis of each component and suggest how they converge within distributed brain networks to support coherent self-voice processing.

Conclusion: This integrative framework aims to advance theoretical and translational work by bridging psychology, neuroscience, clinical research, and voice technology in the context of emerging digital voice environments.

Publicado: 10 de diciembre de 2025

Detection of Induced Psychosis-Like Thinking in User–GenAI Interaction: A Retrieval-Augmented Generation (RAG) Approach

Yi Huang

Artificial intelligence safety, psychosis-like language, delusional thinking detection, large language models, retrieval-augmented generation

Introducción: El rápido crecimiento de la inteligencia artificial generativa ha incrementado la preocupación por su seguridad psicológica. Informes clínicos recientes indican que algunas personas usuarias, especialmente aquellas con mayor vulnerabilidad en salud mental, pueden desarrollar ideas delirantes o experiencias similares a la psicosis durante interacciones con sistemas de IA generativa. Detectar este tipo de lenguaje de riesgo resulta clave para un despliegue seguro de los modelos de lenguaje a gran escala.

Método: El estudio evaluó si los modelos de lenguaje pueden identificar enunciados delirantes y si el uso de *retrieval-augmented generation* (RAG) mejora su rendimiento. Se creó un conjunto de datos de 200 afirmaciones simuladas de usuarios, equilibrado entre afirmaciones delirantes y no delirantes, basado en criterios del DSM-5 y validado por una psicoterapeuta colegiada. El conjunto se dividió en un corpus de recuperación de 100 ejemplos y un conjunto de prueba de 100 ejemplos. Se compararon dos clasificadores basados en GPT-4o-mini: un modelo



base sin RAG y un modelo con RAG que recuperaba los tres ejemplos más similares semánticamente. Cuando una afirmación se clasificaba como delirante, el modelo también asignaba un nivel de gravedad de 1 a 3.

Resultados: El modelo sin RAG alcanzó una precisión del 81% y una sensibilidad de 0,66. El modelo con RAG obtuvo una precisión del 94%, con una mejora clara en la sensibilidad (0,89) y una especificidad perfecta. No obstante, ambos modelos tendieron a infraestimar de forma sistemática la gravedad de los enunciados delirantes, siendo esta infraestimación más marcada en el modelo con RAG.

Conclusión: Los resultados muestran que la incorporación de RAG puede mejorar la detección temprana de lenguaje similar al psicótico en interacciones con sistemas de IA generativa. Sin embargo, la estimación fiable de la gravedad sigue siendo limitada, lo que pone de relieve tanto el potencial como las restricciones actuales de los modelos de lenguaje para la detección de riesgos cognitivos.

Background: With the rapid growth of generative artificial intelligence (GenAI), concerns about its psychological safety have increased. Recent psychiatric reports show that some users—especially those with vulnerable mental health—may develop delusional thoughts or psychosis-like experiences during interactions with GenAI.

Methods: This study examined whether LLMs can identify delusional statements and whether retrieval-augmented generation (RAG) can improve performance. A dataset of 200 simulated user statements (50% delusional, 50% non-delusional) was created based on DSM-5 criteria and validated by a licensed psychotherapist. The dataset was divided into a 100-item retrieval corpus and a 100-item test set. Two GPT-4o-mini classifiers were evaluated: a baseline non-RAG model and a RAG-enhanced model that retrieved the three most semantically similar examples from the corpus. When a statement was classified as delusional, models also assigned a severity rating (1–3).

Results: The non-RAG model achieved 81% accuracy with a recall of 0.66, whereas the RAG model reached 94% accuracy with improved recall (0.89) and perfect specificity. However, both models systematically underestimated severity, with the RAG model showing greater underestimation.

Conclusion: These findings indicate that RAG can enhance early detection of psychosis-like language but is insufficient for reliable severity assessment, highlighting both the potential and the limits of current LLM-based cognitive-risk detection.

Publicado: 10 de diciembre de 2025



Linking interpersonal childhood adversity to mental health: A scoping review of trust, mentalizing, agency and interpersonal emotion regulation as candidate social-transactional mechanisms

Ritika Chokhani, Alex Lloyd, Essi Viding, Mattia Gerin, Eamon McCrory

Childhood interpersonal adversity, trust, mentalizing, agency, interpersonal emotion regulation, mental health

Introducción: El modelo neurocognitivo social-transaccional plantea que la adversidad interpersonal en la infancia incrementa el riesgo de problemas de salud mental al modificar la forma en que las personas se relacionan con su entorno social. Este trabajo tuvo como objetivo mapear la evidencia existente sobre cuatro mecanismos candidatos —confianza, mentalización, agencia y regulación emocional interpersonal— que podrían vincular la adversidad interpersonal temprana con resultados en salud mental.

Método: Se realizó una revisión exploratoria mediante búsquedas sistemáticas en cuatro bases de datos. Se incluyeron estudios cuantitativos que evaluaran adversidad interpersonal en la infancia, al menos uno de los mecanismos candidatos y resultados en salud mental dentro de la misma muestra. Un segundo revisor evaluó de forma independiente el 10% de los registros en cada fase del proceso de cribado.

Resultados: Un total de 78 estudios cumplieron los criterios de inclusión. En el caso de la confianza (15 estudios), la adversidad interpersonal infantil se asoció con mayor desconfianza y credulidad, que a su vez predijeron síntomas internalizantes, externalizantes y de trastornos de la personalidad. En mentalización (37 estudios), se observó una asociación con mayor incertidumbre y menor precisión en la comprensión de estados mentales, consistentemente vinculada con síntomas internalizantes, externalizantes, psicóticos y de trastornos de la personalidad. En agencia (26 estudios), la adversidad se relacionó con un locus de control más externo, asociado con síntomas internalizantes y psicóticos. La evidencia sobre regulación emocional interpersonal fue limitada (3 estudios), aunque sugirió posibles vínculos con síntomas internalizantes. En conjunto, la mayoría de los hallazgos procedían de estudios transversales y autoinformados, con escasa investigación longitudinal o basada en tareas.

Conclusión: La confianza, la mentalización y la agencia emergen como mecanismos sociales-transaccionales transdiagnósticos prometedores a través de los cuales la adversidad interpersonal en la infancia influye en la salud mental, mientras que la evidencia para la regulación emocional interpersonal es más limitada. A pesar de las lagunas metodológicas, estos mecanismos podrían constituir objetivos abordables para la investigación en intervenciones y para la práctica clínica informada en trauma.



Objectives: The neurocognitive social-transactional model posits that interpersonal adversity during childhood heightens risk for mental health difficulties by altering how individuals engage with their social world. We conducted a scoping review to map current evidence for four candidate mechanisms, Trust, Mentalizing, Agency and Interpersonal Emotion Regulation (IER), linking interpersonal childhood adversity to mental health outcomes.

Methods: We systematically searched four databases and identified quantitative studies that measured interpersonal childhood adversity, one or more of the candidate mechanisms, and mental health outcomes in the same sample. A second reviewer independently assessed 10% of records at each screening stage.

Results: Seventy-eight studies met inclusion criteria. For Trust (15 studies), interpersonal childhood adversity was associated with greater distrust and credulity which in turn predicted internalizing, externalizing and personality disorder symptoms. For Mentalizing (37 studies), it was associated with heightened uncertainty and reduced accuracy in understanding mental states, which was consistently linked with internalizing, externalizing, psychosis and personality disorder symptoms. For Agency (26 studies), it was associated with a more external locus of control, which was associated with internalizing and psychosis symptoms. Evidence for IER was limited (3 studies) but suggested potential links to internalizing symptoms. Across mechanisms, findings were largely cross-sectional and based on self-report, with few longitudinal or task-based studies.

Conclusion: Trust, Mentalizing and Agency emerged as promising transdiagnostic social-transactional mechanisms through which interpersonal childhood adversity shapes mental health outcomes, with more limited evidence for IER. Despite methodological gaps, these mechanisms may represent tractable targets for intervention research and trauma-informed clinical practice.

Publicado: 27 de noviembre de 2025



Willingness to trust is reduced by loneliness and paranoia

Gabriele Bellucci, Mehdi Keramati, Esther Hanssen, Anne-Kathrin Fett

Loneliness, paranoia, social decision-making, trust, reciprocity, psychosis

Introducción: La soledad se asocia con conductas sociales negativas que deterioran las relaciones interpersonales. Sin embargo, los mecanismos subyacentes a esta asociación siguen siendo poco comprendidos. Este estudio examinó la relación entre pensamientos paranoides y la disposición de personas solas a basarse en expectativas de reciprocidad de la pareja en una tarea de interacción social.

Método: Se analizó el comportamiento de 54 participantes, con y sin psicosis, en un juego de inversión diseñado para evaluar la confianza y la respuesta a la reciprocidad de la pareja. Se midieron los niveles de soledad y paranoia, así como la conducta de confianza tras rupturas de confianza y la sensibilidad a los cambios en la reciprocidad del compañero. Se aplicaron modelos computacionales para estimar el grado en que las personas se apoyaban en expectativas de reciprocidad.

Resultados: La soledad y la paranoia mostraron una fuerte correlación entre sí y se asociaron con una conducta más desconfiada tras las rupturas de confianza. Las personas más solas y más paranoides mostraron mayor sensibilidad a los cambios en la reciprocidad de la pareja. Además, las personas con mayor soledad confiaron menos incluso en parejas altamente recíprocas. El modelado computacional reveló que los individuos más solitarios y paranoides eran menos propensos a basarse en expectativas de reciprocidad del otro. Estos efectos se observaron tanto en pacientes como en participantes control.

Conclusión: Los hallazgos muestran que la soledad se relaciona con alteraciones en las expectativas sociales y en la conducta de confianza, independientemente de la presencia de psicosis. La soledad y la paranoia desempeñan un papel relevante tanto en poblaciones clínicas como generales, con implicaciones importantes para la calidad y el mantenimiento de las relaciones interpersonales.

Background: Loneliness is associated with negative social behaviors impairing social relationships. However, the underlying mechanisms are poorly understood.

Methods: We investigated the relationship between paranoid thoughts and lonely individuals' willingness to rely on expectations of partner reciprocity in an investment game with individuals with and without psychosis (54 participants).

Results: We found that loneliness and paranoia were strongly correlated with each other and with more distrustful behavior after breaches of trust. Sensitivity to changes in partner reciprocity were higher in lonelier and more paranoid individuals. Lonelier individuals also trusted highly-reciprocating partners less.



Computational modeling revealed that lonelier and more paranoid individuals were less willing to rely on expectations of partner reciprocity. Importantly, these effects were observed in both patients and controls.

Conclusion: These findings demonstrate how loneliness relates to social behaviors and expectations, pointing to important downstream implications for lonely individuals' relationships.

Publicado: 20 de noviembre de 2025

Social networks and symptomatology in recently hospitalized individuals with schizophrenia-spectrum disorders: A six-month longitudinal study

Gabriella V. Schock, Andrew R. Kittleson, Annalise S. Halverson, Jinyuan Liu, Julia M. Sheffield

Social networks, schizophrenia-spectrum disorders, paranoia, positive symptoms, negative symptoms, hospitalization

Introducción: Las personas con psicosis suelen presentar redes sociales reducidas. Sin embargo, se conoce poco sobre cómo evolucionan estas redes tras un episodio agudo que requiere hospitalización y sobre su relación longitudinal con la gravedad de los síntomas, incluida la paranoia, así como con la duración de la enfermedad.

Método: Se analizaron datos longitudinales de personas con trastornos del espectro de la esquizofrenia recientemente hospitalizadas (SSD; N = 68) y de un grupo no clínico (NC; N = 70) durante los seis meses posteriores al alta. Mediante modelos lineales mixtos ajustados por edad y sexo, se evaluó el compromiso con la red social autoinformada, la gravedad de los síntomas positivos y negativos, y los niveles de paranoia. También se examinó el impacto del estadio de la enfermedad en la evolución de las redes sociales dentro del grupo SSD.

Resultados: Las personas con SSD informaron redes sociales significativamente más reducidas que el grupo no clínico en todos los momentos de evaluación. La alteración fue mayor en las relaciones de amistad que en los vínculos familiares. Las redes sociales se mantuvieron estables a lo largo del tiempo en ambos grupos, independientemente del estadio de la enfermedad en el grupo clínico. Un menor compromiso social con la familia al inicio predijo aumentos posteriores en la paranoia. A la inversa, una mayor paranoia basal predijo una reducción más marcada de la red de amistades con el tiempo. Además, el tamaño y la calidad de la red social se asociaron con la gravedad de los síntomas negativos tanto al inicio como de forma longitudinal.



Conclusión: Los resultados sugieren que la gravedad clínica influye en las alteraciones de las redes sociales observadas en los trastornos psicóticos, especialmente durante periodos vulnerables de recuperación tras una hospitalización. Estos hallazgos subrayan la necesidad de intervenciones específicas tras el alta que aborden los déficits en las redes sociales, en particular en el ámbito familiar y en personas con paranoia intensa.

Background: Individuals with psychosis have reduced social networks; however, little is known about whether social networks change over time, particularly after an acute episode requiring hospitalization. Furthermore, longitudinal associations between social network engagement, symptom severity, and the impact of illness duration remains unclear.

Methods: Using linear mixed models controlling for age and sex, we analyzed self-reported social network engagement, positive and negative symptom severity, and self-reported paranoia in individuals with a schizophrenia-spectrum disorder (SSD; N = 68) during six months following psychiatric hospitalization and compared them with non-clinical participants (NC; N = 70).

Results: SSD participants reported significantly lower social network scores than NC participants at every time point, and significantly greater disruption in friendship than family ties. Social networks remained stable over time in both groups, regardless of illness stage in the SSD group. Lower social engagement with family at baseline predicted greater increases in paranoia over time. Conversely, more severe paranoia at baseline predicted greater reductions in friend networks over time. Social network engagement was also related to negative symptom severity at baseline and longitudinally.

Conclusion: These findings suggest that clinical severity may influence the social network disruptions commonly found in psychotic disorders, especially during vulnerable periods of recovery following a psychotic exacerbation. The results highlight the need for tailored post-discharge interventions that address social network deficits, particularly within families and in the context of severe paranoia.

Publicado: 14 de noviembre de 2025



Manuscritos en MedRxiv y BioRxiv

Se trata de dos plataformas en línea y de acceso gratuito para archivar y difundir manuscritos inéditos (*preprints*) en el ámbito de las ciencias médicas, clínicas y biológicas.

- **MedRxiv:** fundado por Cold Spring Harbor Laboratory (CSHL), una institución educativa y de investigación sin ánimo de lucro, la Universidad de Yale y BMJ, un proveedor mundial de conocimientos sobre atención sanitaria.
<https://www.medrxiv.org/>
- **BioRxiv:** gestionado por Cold Spring Harbor Laboratory (CSHL), una institución educativa y de investigación sin ánimo de lucro.
<https://www.biorxiv.org/>

Association between bullying, state-level policy and mental health symptoms in gender diverse youth

Dylan E. Hughes, Sarah L. Zapetis, Arianna Mordy, Daisy Lopez, Vanessa Calderon, Laura Adery, Rachel Martino, Sarah E. Chang, Lucina Q. Uddin, Carlos Cardenas-Iniguez, Richard T. Lebeau, Natalia Ramos, Lauren C. Ng, Katherine H. Karlsgodt, Carrie E. Bearden

bullying victimization, gender diverse youth, psychotic-like experiences, state-level policy, stigma, mental health

Importancia: A medida que aumenta el porcentaje de jóvenes en Estados Unidos que se identifican como transgénero y con diversidad de género, más jóvenes se enfrentan a estigmatización y discriminación sociales y estructurales vinculadas a la identidad. Se sabe poco sobre el impacto del estigma en las experiencias similares a la psicosis en jóvenes con diversidad de género.

Objetivo: Evaluar el impacto de la victimización por acoso escolar y de las políticas a nivel estatal sobre las experiencias similares a la psicosis en jóvenes con diversidad de género.

Diseño: En este estudio de cohorte prospectivo se realizaron análisis transversales y longitudinales utilizando datos recogidos entre 2017 y 2022 en 17 estados.

Ámbito: El estudio Adolescent Brain Cognitive Development (ABCD) es un estudio longitudinal poblacional en Estados Unidos que sigue y caracteriza en profundidad a adolescentes desde los 9 hasta los 18 años.

Participantes: Los análisis transversales incluyeron datos de 9.112 participantes (edad media = 13 ± 0,6) recogidos entre 2019 y 2022. Los análisis longitudinales incluyeron 4.529 participantes con datos recogidos a lo largo de cinco olas entre 2017 y 2022.



Exposiciones: Frecuencia autoinformada de victimización por acoso escolar y datos sobre políticas estatales anuales relacionadas con la identidad de género.

Resultados principales: Experiencias similares a la psicosis autoinformadas y el malestar asociado, medidos mediante el Prodromal Questionnaire – Brief Child Version.

Resultados: A partir de una medida dimensional de género, se identificaron 689 adolescentes como los más diversos en género (es decir, menos congruentes con el sexo asignado al nacer) y 8.240 como los menos diversos en género (es decir, más congruentes con el sexo asignado al nacer). Las tasas de victimización por acoso escolar y de experiencias similares a la psicosis fueron significativamente más elevadas en el grupo con mayor diversidad de género en comparación con el grupo con menor diversidad de género, y el acoso explicó parcialmente la diferencia en experiencias similares a la psicosis (efecto indirecto = 0,11, $p < 2 \times 10^{-16}$; efecto directo = 0,52, $p < 1 \times 10^{-16}$). Los adolescentes con diversidad de género mostraron una mayor sensibilidad a los efectos del acoso sobre las experiencias similares a la psicosis (interacción $b = 0,14$; IC 95% [0,09, 0,19]; $p = 8,5 \times 10^{-8}$). Además, la persistencia de legislación no favorable durante cuatro años se asoció con incrementos significativamente mayores de experiencias similares a la psicosis a lo largo del tiempo en jóvenes con diversidad de género (interacción $b = 0,30$; IC 95% [0,20, 0,40]; $p = 2,2 \times 10^{-8}$).

Conclusiones: Estos hallazgos indican que la victimización por acoso escolar y la legislación no favorable pueden explicar tasas más elevadas y crecientes de experiencias similares a la psicosis en jóvenes con diversidad de género.

Importance: As the percentage of young people in the United States identifying as transgender and gender diverse increases, more youth face identity-linked social and structural stigma and discrimination. Little is known about the impact of stigma on psychotic-like experiences in gender diverse youth.

Objective: To assess the impact of bullying victimization and state-level policies on psychotic-like experiences among gender diverse youth.

Design: In this prospective cohort study, cross-sectional and longitudinal analyses were conducted using data collected between 2017 and 2022 across 17 states.

Setting: The Adolescent Brain Cognitive Development (ABCD) Study is a U.S. population-based longitudinal study that follows adolescents from age 9 to 18.

Participants: Cross-sectional analyses included 9,112 participants. Longitudinal analyses included 4,529 participants with data across five waves.



Exposures: Self-reported bullying victimization and annual state-level policies related to gender identity.

Main outcomes: Psychotic-like experiences and associated distress measured by the Prodromal Questionnaire – Brief Child Version.

Results: Rates of bullying and psychotic-like experiences were higher in gender diverse youth. Bullying partially mediated group differences, and gender diverse youth showed greater sensitivity to bullying effects. Persistent unsupportive legislation was associated with greater increases in psychotic-like experiences over time.

Conclusions: Bullying victimization and unsupportive state-level policies may contribute to elevated and increasing psychotic-like experiences in gender diverse youth.

Publicado: 27 de diciembre de 2025



STELLA: Safety Testing Engine for Large Language Assistants

Roy H. Perlis, Annas Bin Adil, Kit Dobyns

Large language models, safety evaluation, mental health, chatbots, simulation-based testing

Introducción: Los asistentes basados en grandes modelos de lenguaje se utilizan cada vez más en el ámbito sanitario y ofrecen un potencial relevante para ampliar el acceso a la atención. No obstante, crece la preocupación por el riesgo de que estos sistemas respondan de forma inadecuada ante personas en crisis o puedan afectar negativamente a la salud mental en determinadas circunstancias.

Método: Los autores desarrollaron e implementaron un sistema automatizado para evaluar las respuestas de asistentes de IA, tanto por voz como por texto, ante distintos escenarios de salud. El sistema incorpora usuarios simulados con características definidas, escenarios de interacción en múltiples turnos y diseños que permiten comparar cohortes, incluidos ensayos aleatorizados simulados generados mediante instrucciones en lenguaje natural. Las transcripciones de las sesiones se evalúan en términos de seguridad, eficacia y compromiso del usuario mediante rúbricas y ejemplos predefinidos, utilizando un conjunto de modelos de lenguaje como jueces, lo que permite señalar intercambios específicos para revisión manual. Para ilustrar el enfoque, se evaluaron 10 escenarios de seguridad en 11 chatbots de modelos avanzados, con cinco perfiles de usuario seguidos durante 10 intercambios, y un subconjunto evaluado con cinco perfiles adicionales.

Resultados: La proporción total de respuestas marcadas como potencialmente dañinas osciló entre el 3,2% en ChatGPT-5.2 y el 34,0% en Grok-4.1-fast-non-reasoning. La proporción de respuestas señaladas por no ofrecer contenido beneficioso varió entre el 19,6% en ChatGPT-5.2 y el 66,0% en Grok-4.1-fast-reasoning. En conjunto, el contenido inseguro tendió a aumentar a lo largo de los turnos de interacción, mientras que los fallos en proporcionar contenido beneficioso se incrementaron en torno a un 0,7% por turno.

Conclusión: Un entorno de pruebas basado en simulación permite caracterizar y comparar de forma rápida el rendimiento de asistentes basados en modelos de lenguaje según criterios estandarizados. Los modelos evaluados muestran una variabilidad considerable en métricas de seguridad y utilidad. Este tipo de estrategias puede acelerar los esfuerzos para garantizar que los chatbots aporten beneficios y minimicen riesgos cuando se aplican a la salud mental y el bienestar.

Background: Assistants incorporating large language models are increasingly applied in the context of health care, where they represent a promising means of expanding access to care. However, there is growing recognition of the risks that these chatbots may fail to respond appropriately to individuals in crisis, and may adversely affect mental health in some circumstances.

Methods: We developed and implemented an automated system for assessing voice



or text AI assistant response to users across a range of health scenarios. This set of tools incorporates simulated users with a specified set of characteristics; scenarios in which they interact with a chatbot over multiple rounds; and designs that allow multiple cohorts to be compared. Study designs including simulated randomized trials can be generated via natural language prompts. Chatbot session transcripts are then quantified in terms of safety, efficacy, and user engagement according to prespecified rubrics and exemplars with an ensemble of judging language models, allowing specific exchanges to be flagged for manual review. To illustrate this approach, we assessed 10 safety scenarios in 11 frontier language model chatbots, including Claude Opus 4.5, ChatGPT-5.2, and Gemini 3, using 5 personas, each followed over 10 exchanges, with a subset assessed for an additional 5 personas.

Results: Total proportion of responses flagged for possible harmful content ranged from 3.2% (95% CI 2.0–5.1%) for GPT-5.2 to 34.0% (95% CI 30.0–38.3%) for Grok-4.1-fast-non-reasoning. Total proportion of responses flagged for failing to provide beneficial content ranged from 19.6% (95% CI 16.4–23.3%) for GPT-5.2 to 66.0% (95% CI 61.7–70.0%) for Grok-4.1-fast-reasoning. In aggregate, proportion of unsafe content increased across turns; for failure to provide beneficial content, by 0.7% per turn (95% CI 0.3%–1.1%).

Conclusion: A simulation-based test harness can facilitate the rapid characterization and comparison of large language model assistant performance according to standardized rubrics. Existing frontier models vary substantially on these metrics. Simulation strategies such as this one may accelerate efforts to ensure that chatbots yield benefit rather than harm to users who seek to apply them to address mental health and well-being.

Publicado: 27 de diciembre de 2025



Development and Evaluation of Machine Learning Models to Predict Mechanical Restraint and Related Coercive Measures in Hospital Psychiatry

Sara Kolding, Jakob Grøhn Damgaard, Martin Bernstorff, Lasse Hansen, Søren Dinesen Østergaard, Andreas Aalkjær Danielsen

Machine learning, mechanical restraint, coercive measures, psychiatric hospitalization, electronic health records

Introducción: El uso de medidas coercitivas en hospitales psiquiátricos plantea importantes desafíos clínicos y éticos. Con el objetivo de apoyar estrategias preventivas, este estudio desarrolló y evaluó modelos de aprendizaje automático para predecir la aplicación de sujeción mecánica y un resultado compuesto que incluye otras medidas coercitivas relacionadas.

Método: Se utilizaron registros electrónicos de salud de personas adultas con al menos un ingreso en los Servicios Psiquiátricos de la Región Central de Dinamarca entre 2015 y 2021. Para cada día de hospitalización, se entrenaron modelos XGBoost para predecir la probabilidad de sujeción mecánica o de sujeción compuesta (mecánica, química o manual) en las 48 horas siguientes. Los hiperparámetros se optimizaron según el área bajo la curva ROC mediante validación cruzada de cinco pliegues en el 85% de los datos, y el rendimiento se evaluó en un conjunto de prueba independiente del 15%.

Resultados: La cohorte incluyó 16.834 pacientes, con 45.179 estancias hospitalarias y 687.388 días de predicción. De estos, 2.736 días fueron seguidos por algún episodio de sujeción en las 48 horas posteriores, incluidos 983 episodios de sujeción mecánica. Los modelos finales incorporaron 2.389 predictores derivados de variables demográficas, diagnósticos, medicación y notas clínicas. El modelo de sujeción mecánica alcanzó un AUROC de 0,921 (IC 95%: 0,918–0,922) y un valor predictivo positivo del 4,9% al clasificar como positivos el 1% de los días con mayor riesgo. El modelo compuesto obtuvo un AUROC de 0,912 (IC 95%: 0,909–0,913) y un valor predictivo positivo del 4,2% al predecir sujeción mecánica, y un AUROC de 0,900 (IC 95%: 0,898–0,900) con un valor predictivo positivo del 10,4% al predecir sujeción compuesta.

Conclusión: Los resultados indican que incluir medidas coercitivas relacionadas en el entrenamiento del modelo no mejora la capacidad discriminativa para predecir sujeción mecánica, pero sí aumenta el valor predictivo positivo al predecir resultados compuestos, debido a su mayor prevalencia. Estos hallazgos subrayan la importancia de cómo se formula el problema en los modelos predictivos clínicos y señalan la necesidad de validaciones externas en distintos contextos temporales, geográficos y demográficos.

Introduction: Use of coercive measures in psychiatric hospitals is clinically and ethically challenging. Aiming to support prevention, we developed and evaluated machine learning models to predict both mechanical restraint and a broader composite outcome that includes



related coercive measures.

Methods: The dataset comprised electronic health records (EHR) from adults (≥ 18 years) who had at least one admission to the Psychiatric Services in the Central Denmark Region between 2015 and 2021. For each inpatient day, XGBoost machine learning models were trained to predict mechanical restraint or composite (mechanical, chemical, or manual) restraint within 48 hours. Hyperparameters were optimised for the area under the receiver operating characteristic curve (AUROC) using five-fold cross validation on 85% of the data, with performance validated on a held-out 15% test set.

Results: The cohort included 16,834 patients with 45,179 inpatient stays, covering 687,388 prediction days. Of these, 2,736 days were followed by a restraint episode within 48 hours, including 983 episodes of mechanical restraint. The final models were trained on 2,389 EHR-based predictors, derived from demographics, diagnoses, medications, and clinical notes. The mechanical restraint model achieved an AUROC of 0.921 (95% CI: [0.918–0.922]) and a positive predictive value (PPV) of 4.9% when classifying the top 1% of risk scores as positive. The composite model achieved an AUROC of 0.912 (95% CI: [0.909–0.913]) and a PPV of 4.2% when predicting mechanical restraint, and 0.900 (95% CI: [0.898–0.900]) with a PPV of 10.4% when predicting composite restraint.

Conclusion: The results indicate that incorporating related coercive measures into model training did not improve discrimination for predicting mechanical restraint but did increase PPV when predicting composite restraint, highlighting the importance of outcome framing in prediction of rare clinical events.

Publicado: 16 de diciembre de 2025

A transdiagnostic evaluation of adult temperamental profiles and their interactions with childhood adversities across major psychiatric disorders

Pavithra Dayalamurthy, Srinivas Balachander, Sai Priya Lakkireddy, Mahashweta Bhattacharya, Mino Susan Joseph, Pramod Kumar, Anand Jose Kannampuzha, Sreenivasulu Mallappagari, Shruthi Narayana, Alen Chandy Alexander, Moorthy Muthukumaran, Joan C. Puzhakkal, Navya Spurthi Thatikonda, Sweta Sheth, Furkhan Ali, Sowmya Selvaraj, Vanteemar S. Sreeraj, Bharath Holla, Ganesan Venkatasubramanian, John P. John, Biju Viswanath, cVEDA Consortium, ADBS-CBM Consortium, Vivek Benegal, YC Janardhan Reddy, Sanjeev Jain

Temperament, adverse childhood experiences, transdiagnostic psychiatry, schizophrenia, bipolar disorder, obsessive-compulsive disorder, alcohol use disorder

Antecedentes: Los perfiles temperamentales y las experiencias adversas en la infancia se han descrito en la mayoría de los principales trastornos psiquiátricos. Sin embargo, la naturaleza de esta relación es compleja y ha sido poco estudiada.

Métodos: Se incluyeron un total de 1.162 personas (N = 532 afectadas; N = 457 familiares de primer grado no afectados y N = 173 controles sanos). Las personas



afectadas y sus familiares de primer grado procedían de 393 familias multiplex con esquizofrenia, trastorno bipolar, trastorno obsesivo-compulsivo o trastorno por consumo de alcohol. Se evaluaron mediante el Adult Temperament Questionnaire autoinformado y el cuestionario internacional de experiencias adversas en la infancia. Se aplicaron modelos de regresión de efectos mixtos, tanto frecuentistas como bayesianos, para analizar las asociaciones entre los dominios temperamentales, la puntuación de experiencias adversas en la infancia y su interacción con cada diagnóstico o riesgo familiar. La validez externa del modelo final se evaluó para predecir el diagnóstico en una cohorte independiente. Se utilizaron curvas ROC y el área bajo la curva para valorar la precisión predictiva.

Resultados: Todos los dominios temperamentales —afectividad negativa, extraversión, control esforzado y sensibilidad de orientación— mostraron asociaciones significativas en configuraciones específicas para cada uno de los cuatro diagnósticos y su riesgo familiar. Las experiencias adversas en la infancia y sus interacciones se asociaron de forma significativa únicamente con el trastorno por consumo de alcohol. En el conjunto de validación externa, el modelo final predijo con éxito el trastorno por consumo de alcohol (AUC = 0,82), el trastorno bipolar (AUC = 0,75) y el trastorno obsesivo-compulsivo (AUC = 0,74), pero no la esquizofrenia (AUC = 0,53).

Conclusiones: Estos hallazgos aportan información sobre la complejidad del temperamento en los principales trastornos psiquiátricos y subrayan la importancia de evaluar el temperamento junto con la adversidad infantil para caracterizar el riesgo.

Background: Temperamental profiles and adverse childhood experiences (ACEs) have been reported in most major psychiatric disorders. However, the nature of this relationship is complex and under-studied.

Methods: A total of 1162 individuals (N = 532 affected; N = 457 unaffected first-degree relatives and N = 173 healthy controls) were included. Affected individuals and their first-degree relatives were from 393 multiplex families with schizophrenia, bipolar disorder, obsessive-compulsive disorder, or alcohol use disorder. Participants were assessed using the self-reported Adult Temperament Questionnaire and the ACEs International Questionnaire. Frequentist and Bayesian mixed-effects regression models were used to examine associations between temperament domains, ACEs scores, and their interaction with diagnosis or familial risk. External validity of the final model was tested in an independent cohort. Receiver operating characteristic curves and area under the curve were used to assess predictive accuracy.

Results: All temperamental domains—negative affectivity, extraversion, effortful control, and orienting sensitivity—showed significant associations in diagnosis- and risk-specific configurations. ACEs and their interactions were significantly associated only with alcohol use disorder. In external validation, the model successfully



predicted alcohol use disorder (AUC = 0.82), bipolar disorder (AUC = 0.75), and obsessive-compulsive disorder (AUC = 0.74), but not schizophrenia (AUC = 0.53).

Conclusions: These findings provide unique insights into the complexity of temperament across major psychiatric disorders and highlight the importance of considering temperament together with childhood adversity in risk characterization.

Publicado: 16 de diciembre de 2025

TRUSTING: An International Multicenter Observational Study of Speech-Based Relapse Prediction in Psychosis Using Explainable AI

Roya M. Hüppi, Lucía Bautista, Giacomo Cecere, Sandra A. Just, Sanne Koops, Musarrat Hussain, Enrico Tedeschi, Stephan Benke-Bruderer, Emre Bora, John Lyne, Stefan Kaiser, Elodie Sprüngli-TOffel, Matthias Kirschner, Karl Øyvind Mikalsen, Lars Ailo Bongo, Erik Van der Eycken, Filip Španiel, Brita Ellevåg, Iris E. Sommer, Wolfram Hinzen, Philipp Homan

Psychotic relapse prediction, speech-based biomarkers, artificial intelligence monitoring, schizophrenia, digital phenotyping

Introducción: El curso de los trastornos psicóticos suele caracterizarse por recaídas. Las señales tempranas de recaída varían entre personas y resultan difíciles de detectar en la práctica clínica, especialmente en contextos ambulatorios. El habla ofrece un marcador clínico cuantitativo con potencial para identificar estas señales de alerta precoz. El proyecto europeo Horizon TRUSTING tiene como objetivo desarrollar y evaluar un sistema de monitorización basado en el habla para predecir recaídas psicóticas inminentes, así como analizar su viabilidad y usabilidad.

Método: Se trata de un estudio observacional multicéntrico que incluirá a 240 personas adultas con trastornos psicóticos en remisión y con riesgo de recaída, junto con un grupo comparativo de 120 personas sanas emparejadas por edad y sexo. El estudio se llevará a cabo en seis centros y en seis idiomas distintos: alemán, francés, neerlandés, inglés, checo y turco. El periodo de seguimiento será de seis meses. A través de la aplicación móvil TRUSTING se recopilarán semanalmente grabaciones de voz mediante tareas de habla, información sobre adherencia al tratamiento, consumo de sustancias, estado de ánimo, ansiedad y calidad del sueño, así como datos motores a partir de una tarea de golpeteo.

Análisis: Los criterios principales de evaluación incluyen el rendimiento del modelo para la predicción de recaídas, la adherencia de las personas usuarias, la calidad de las transcripciones, la usabilidad de las grabaciones y la usabilidad global del sistema. El análisis principal de adherencia, calidad de transcripción y usabilidad consistirá en una descripción no ajustada de proporciones con intervalos de confianza de Wilson al 95%. La capacidad predictiva del sistema para la ocurrencia de recaídas se evaluará mediante el área bajo la curva ROC. De forma exploratoria, se analizarán posibles



marcadores del habla asociados al riesgo de recaída.

Consideraciones éticas y difusión: El estudio cuenta con la aprobación del comité ético swissethics. Los resultados se difundirán a través de publicaciones en revistas científicas revisadas por pares, congresos especializados y actividades públicas relacionadas con la salud mental.

Introduction: The course of psychotic disorders typically involves relapses. Early warning signs vary between individuals and are difficult to detect in clinical practice, especially in outpatient settings. Speech provides a quantitative clinical marker for detecting such early warning signs. The EU Horizon project TRUSTING aims to develop and evaluate a speech-based monitoring system for predicting imminent psychotic relapses. The study will examine the potential for prospective relapse prediction, and feasibility and usability of the monitoring system.

Methods and analysis: In this multicenter observational study, $n = 240$ remitted and at-risk-of-relapse adults with psychotic disorders and a comparison group with $n = 120$ healthy participants matched by age and sex will be examined at six sites and in six different languages. The follow-up period is 6 months. The TRUSTING smartphone app will be used to collect weekly voice recordings through speech tasks; information on medication adherence, substance use, mood, anxiety, and sleep quality; and motor data from a tapping task. Primary endpoints encompass model performance for relapse prediction, user adherence, transcription quality, usability of recordings, and overall system usability. User adherence and usability outcomes will be described using unadjusted proportions with 95% Wilson confidence intervals. Predictive value for relapse occurrence will be assessed using the area under the receiver operating characteristic curve. Exploratory analyses will examine speech-based markers associated with relapse risk.

Ethics and dissemination: This study has been approved by swissethics. Findings will be disseminated through peer-reviewed publications, scientific conferences, and public mental health events.

Publicado: 15 de diciembre de 2025



Beyond 'Expressed Emotion': Development and validation of a contextual measure of family communication and involvement in the context of psychosis in Ethiopia

Dessalegn Kiross, Denise Taylor, Kassahun Habtamu, Wubalem Fekadu, Atalay Alem, Mekonnen Tsehay, Eleni Misganaw, Crick Lund, Tessa Roberts, Ashok Malla, Alex Cohen, Kia-Chong Chua, Tosin Popoola, Charlotte Hanlon

Family communication, family involvement, psychosis, cultural adaptation, scale development, Ethiopia

Antecedentes y objetivo: Los entornos familiares de cuidado influyen de forma significativa en los resultados de la psicosis, pero las medidas existentes, como *Expressed Emotion*, carecen de relevancia cultural en contextos no occidentales. Este estudio tuvo como objetivo desarrollar y validar en Etiopía una escala de Comunicación e Implicación Familiar (*Family Communication and Involvement*, FCI) socio-culturalmente apropiada.

Métodos: Se generó una escala FCI inicial (versión 0) integrando ítems procedentes de medidas existentes sobre comunicación familiar con hallazgos de un estudio etnográfico realizado con familias etíopes que cuidaban a un familiar con psicosis. Para garantizar la validez de contenido y la claridad, los ítems se sometieron a traducción, retrotraducción, revisión por consenso de expertos y entrevistas cognitivas ($n = 20$). La FCI v1.0 resultante (37 ítems) se pilotó en 201 personas con psicosis en zonas rurales de Etiopía. El análisis factorial exploratorio, las correlaciones ítem-ítem e ítem-total y la fiabilidad test-retest ($n = 50$) informaron discusiones con actores clave para reducir el número de ítems, lo que dio lugar a una estructura bifactorial de 24 ítems. La escala final de 24 ítems (FCI v2) se validó en una muestra independiente de 401 personas con psicosis en Addis Abeba. El análisis factorial confirmatorio evaluó la validez estructural y la validez convergente se examinó frente a la *WHO Disability Assessment Schedule* y la *Brief Psychiatric Rating Scale–Expanded*.

Resultados: La escala mostró un rendimiento psicométrico sólido, con excelente consistencia interna (alfa de Cronbach = 0,92) y validez convergente aceptable, correlacionando con discapacidad (WHODAS, $r = -0,45$) y gravedad sintomática (BPRS-E, $r = -0,48$). El análisis factorial confirmatorio respaldó la estructura bifactorial de 24 ítems con un ajuste del modelo aceptable.

Conclusión: La escala FCI es una herramienta culturalmente fundamentada y psicométricamente robusta para evaluar la comunicación y la implicación familiar en Etiopía. Presenta un alto potencial de uso en otros contextos no occidentales donde el apoyo familiar y los factores socioculturales influyen en la atención en salud mental.

Background and Aim: Family caregiving environments significantly influence psychosis outcomes, yet existing measures like *Expressed Emotion* lack cultural



relevance in non-Western settings. This study aimed to develop and validate a socio-culturally appropriate Family Communication and Involvement (FCI) scale in Ethiopia.

Methods: An initial FCI scale (version 0) was generated by integrating items from existing measures of family communication with findings from an ethnographic study with Ethiopian families caring for a relative with psychosis. Items underwent translation, back-translation, expert consensus review, and cognitive interviews ($n = 20$). The resulting FCI v1.0 (37 items) was piloted with 201 people with psychosis in rural Ethiopia. Exploratory factor analysis, item correlations, and test–retest reliability ($n = 50$) informed item reduction, yielding a 24-item two-factor structure. The final FCI v2 was validated in a separate sample of 401 people with psychosis in Addis Ababa using confirmatory factor analysis. Convergent validity was assessed against the WHO Disability Assessment Schedule and the Brief Psychiatric Rating Scale–Expanded.

Results: The scale showed excellent internal consistency (Cronbach’s $\alpha = 0.92$) and acceptable convergent validity, correlating with disability (WHODAS, $r = -0.45$) and symptom severity (BPRS-E, $r = -0.48$). Confirmatory factor analysis supported the 24-item bifactor structure with acceptable model fit.

Conclusion: The FCI scale is a culturally grounded and psychometrically sound tool for assessing family communication and involvement in Ethiopia, with potential applicability in other non-Western settings where family and sociocultural factors shape mental health care.

Publicado: 26 de noviembre de 2025



Manuscritos en Preprints.org

Es una plataforma **multidisciplinar** que ofrece manuscritos con el objetivo de que estos puedan recuperarse de forma permanente y se citen en los trabajos.

Este servidor de preprints está subvencionado por MDPI (Multidisciplinary Digital Publishing Institute) en Basilea, Suiza. <https://www.preprints.org/>

Achieving Mood Stabilization Without Mood Stabilizers: A Retrospective Case Observation of an Adolescent with Bipolar-OCD Symptoms Treated via the NMDA-AMPA Axis

Ngo Cheung

Medicine and Pharmacology, Psychiatry and Mental Health

Introducción: Los adolescentes que presentan simultáneamente conductas obsesivo-compulsivas y una labilidad afectiva marcada plantean un reto terapéutico: la recomendación habitual de litio, valproato u otros estabilizadores del estado de ánimo puede conllevar costes metabólicos y cognitivos, mientras que prescindir de ellos implica el riesgo de un viraje maníaco inducido por antidepresivos.

Descripción del caso: Se describe el caso de un varón de 17 años con antecedente de TDAH en la infancia y rituales de comprobación de larga evolución; durante su estancia en un internado en el extranjero desarrolló rumiación intensa, amplias oscilaciones del estado de ánimo e ideación suicida pasiva. El tratamiento con risperidona a dosis bajas no fue eficaz.

Intervención: En lugar de añadir un estabilizador del estado de ánimo convencional, se introdujo una combinación oral glutamatérgica orientada a reproducir una plasticidad similar a la de la ketamina: dextrometorfano como antagonista NMDA, piracetam para potenciar la transmisión AMPA y 10 mg de fluoxetina nocturna para ralentizar el metabolismo del dextrometorfano.

Resultados: En pocas semanas, el estado de ánimo se estabilizó, los pensamientos intrusivos disminuyeron y el funcionamiento escolar mejoró. El paciente se mantuvo estable sin estabilizadores del estado de ánimo a dosis terapéuticas.

Conclusión: Esta experiencia sugiere que una terapia glutamatérgica oral cuidadosamente combinada puede, en adolescentes seleccionados, aportar beneficios tanto antiobsesivos como estabilizadores del estado de ánimo, evitando las cargas asociadas a los tratamientos tradicionales.

Introduction: Adolescents who show both obsessive-compulsive behaviour and pronounced, labile mood present a treatment challenge: the usual recommendation



of lithium, valproate or other mood stabilisers can bring metabolic and cognitive costs, yet withholding them risks an antidepressant-induced switch to mania.

Case description: A 17-year-old boy illustrates an alternative course. He had childhood ADHD and long-standing checking rituals; while at boarding school overseas he developed intense rumination, wide mood swings and passive suicidal thoughts. Low-dose risperidone failed to help.

Intervention: Rather than add a conventional mood stabiliser, we introduced an oral glutamatergic stack aimed at reproducing ketamine-like plasticity: dextromethorphan for NMDA antagonism, piracetam to potentiate AMPA transmission and 10 mg fluoxetine nightly to slow dextromethorphan metabolism.

Results: Within weeks mood steadied, intrusive thoughts abated and function at school improved. The patient remained well without therapeutic-dose mood stabilisers.

Conclusion: This experience suggests that carefully layered oral glutamatergic therapy may, in selected adolescents, deliver both anti-obsessional and mood-stabilising benefits while avoiding the burdens of traditional agents.

Publicado: 10 de diciembre de 2025





GOBIERNO
DE ESPAÑA

MINISTERIO
DE DERECHOS SOCIALES, CONSUMO
Y AGENDA 2030

SECRETARÍA DE ESTADO
DE DERECHOS SOCIALES



Centro de
Referencia
Estatad de
Atención
Psicosocial

www.creap.imserso.es