Psychiatry Research 366 (2026) 117450

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Original article

Algorithmic minds, human bias: How AI-generated images shape mental health stigma

Luca Pingani [a,b,*], Nicole Lanzi [a], Erika De Marco [a], Giulia Ferrazzi [a], Anna Ghirardello [a,b], Mattia Marchi [a,b], Annamaria Nasi [b], Giorgia Polizzotto [a], Giulia Pollice [a], Maria Chiara Trapasso Colli [a], Gian Maria Galeazzi [a,b]

[a] Department of Biomedical, Metabolic and Neuroscience, University of Modena and Reggio Emilia, Reggio Emilia, Italy
[b] Integrated Department of Mental Health and Pathological Addictions, Azienda USL – IRCCS di Reggio Emilia, Reggio Emilia, Italy

ARTICLE INFO

Keywords:
Mental health stigma
Generative AI
Visual representations
Social distance
Stereotype activation

ABSTRACT

Mental health stigma is shaped by complex social and representational processes increasingly influenced by digital and algorithmic environments. Despite the rapid diffusion of generative artificial intelligence (AI), little is known about how AI-generated visual representations may affect stigma-related responses. Because generative AI systems may reproduce and amplify stereotypes embedded in their training data and in the prompts used to guide image generation, they represent an increasingly relevant channel through which visual representations of mental illness are produced and disseminated. This study examined whether AI-generated images of mental disorders influence social distance depending on the use of stereotype-based prompts. Using a within-subject experimental design, 311 participants evaluated images depicting depression, bipolar disorder, and schizophrenia generated with or without explicitly stereotypical prompts. Overall, stereotype-based prompts were associated with slightly lower social distance compared to the no-prompt condition ($d = 0.44$). However, this effect varied significantly across diagnostic categories. No differences emerged for depression, whereas bipolar disorder showed a large decrease in social distance in the prompt condition ($d = 1.18$). In contrast, schizophrenia showed the expected pattern, with higher social distance following stereotype-based prompts ($d = -0.64$). Mental health knowledge was associated with lower stigma overall but predicted greater sensitivity to stereotype-based representations in schizophrenia. Prior exposure to mental health-related experiences was linked to lower stigma but did not moderate experimental effects. These findings suggest that AI-generated representations may differentially shape stigma across disorders, highlighting the need to consider algorithmic environments in stigma research and in the development of digital communication and anti-stigma interventions.

1. Introduction

Mental health stigma remains a pervasive and consequential barrier to recovery, social inclusion, and access to care worldwide. Stigma has been conceptualized as a multifaceted process involving labeling, stereotyping, separation, status loss, and discrimination, all occurring within a broader context of power relations (Fox et al., 2018; Link and Phelan, 2001). These processes are sustained and reinforced through social, cultural, and communicative mechanisms that shape how mental illness is represented and understood. A substantial body of research has shown that public stigma toward mental disorders is driven by specific stereotype domains—such as dangerousness, unpredictability, and personal responsibility—which in turn elicit emotional reactions and behavioral responses, including social distancing and discrimination (Angermeyer and Dietrich, 2006; Corrigan and Watson, 2002;Pingani et al., 2016a).

Among the various channels through which stigma is constructed and maintained, visual representations play a particularly influential role. Images can convey complex information rapidly and often operate at an implicit level, shaping perceptions and attitudes even in the absence of explicit verbal content. Prior research has demonstrated that media portrayals of mental illness—especially those emphasizing violence, unpredictability, or chronicity—contribute to the reinforcement of stigmatizing beliefs and increased desire for social distance

* Corresponding author at: Health Professions of Rehabilitation Sciences Department of Biomedical, Metabolic and Neuroscience, University of Modena and Reggio Emilia, "San Lazzaro" Campus - Via Amendola 2, 41121 Reggio Emilia, Italy
E-mail address: luca.pingani@unimore.it (L. Pingani).

https://doi.org/10.1016/j.psychres.2026.117450
Received 6 May 2026; Received in revised form 5 August 2026; Accepted 5 September 2026
Available online 6 September 2026

(Stuart, 2006; Wahl, 2003). More recent work has extended this perspective by showing that even subtle variations in representation and framing – such as differences in facial expression (e.g., neutral vs. distressed), contextual cues (e.g., isolation vs. social interaction), or visual emphasis on threat-related features — can influence stigma-related responses, highlighting the importance of how mental illness is visually constructed (Pingani et al., 2016a).

At the same time, stigma is not only a matter of attitudes but also of intended behavior. Measures of social distance have been widely used as a proxy for stigma-related behavioral intentions, capturing individuals' willingness to engage with people with mental health conditions in everyday situations. In line with this perspective, recent work has highlighted the importance of examining stigma as a dynamic process shaped by contextual and situational factors, rather than as a fixed set of beliefs (Brown et al., 2016).

In recent years, advances in artificial intelligence (AI), and particularly in generative models, have transformed the production and dissemination of visual imagery. Text-to-image systems are now capable of generating highly realistic and contextually rich images based on short textual prompts, raising important questions about how these systems may shape social representations of complex phenomena, including mental illness. A growing body of research suggests that AI systems are not neutral tools but may reproduce, amplify, or even transform existing social biases embedded in their training data and in the prompts used to guide generation (Bianchi et al., 2022; Birhane et al., 2021; Luccioni et al., 2023).

In parallel, research on mental health stigma has increasingly highlighted the role of media and communication environments in shaping public attitudes and beliefs, as well as in influencing behavioural intentions and broader social responses toward individuals with mental illness. Exposure to specific types of representations can reinforce or challenge stigma-related processes, influencing both cognitive and affective responses, as well as patterns of social interaction and exclusion (Stuart, 2006; Betton et al., 2015; Chan and Yanos, 2018). More recently, scholars have begun to consider how digital and algorithmic ecosystems may act as new "stigma environments," where representations are not only consumed but also dynamically generated and personalized, potentially amplifying existing stereotypes in novel ways (Andalibi et al., 2023; Wang et al., 2026).

At the same time, the relationship between stereotype activation and stigma-related outcomes is not necessarily straightforward. Contemporary models emphasize that stigma is shaped by complex, context-dependent processes, in which the content, valence, and framing of stereotypes interact with individual and situational factors (Fiske et al., 2002; Livingston and Boyd, 2010). In particular, different mental disorders are associated with distinct stereotype profiles and social meanings, which may lead to heterogeneous responses when these representations are activated. For example, schizophrenia is consistently associated with perceptions of dangerousness and unpredictability, whereas depression is more often linked to beliefs about personal responsibility and reduced functioning, and bipolar disorder may evoke more ambivalent or mixed stereotypes, including both negative and positive attributes (Ellison et al., 2015; Wood et al., 2014). These differences suggest that the impact of visual representations—whether human- or AI-generated—may vary substantially across diagnostic categories.

In addition to representational factors, individual-level variables such as mental health knowledge and prior contact with people with mental health conditions play a critical role in shaping stigma-related responses. Greater knowledge has generally been associated with more favorable attitudes, although its effects are not always consistent and may depend on how information is integrated with existing beliefs (Evans-Lacko et al., 2010; Knaak et al., 2017). Similarly, contact with individuals with mental illness has been identified as one of the most effective strategies for reducing stigma, particularly when it involves meaningful and positive interactions (Corrigan et al., 2012; Morgan et al., 2018). However, it remains unclear how these factors interact with exposure to AI-generated visual representations.

Despite the growing use of artificial intelligence (AI) in generating visual content, little is known about how AI-generated representations may reproduce or amplify existing stereotypes associated with mental illness. Given that visual imagery plays a crucial role in shaping public attitudes, understanding the potential impact of AI-generated depictions on stigma-related outcomes is of increasing relevance.

The present study aimed to examine whether AI-generated images of mental disorders differ in their impact on social distance depending on the presence of explicitly stereotypical prompts. Specifically, we investigated whether images generated using stereotype-based prompts elicit higher levels of desired social distance compared to images generated without such prompts. In addition, the study explored whether this effect varies across different diagnostic categories commonly included within severe mental illness—namely depression, bipolar disorder, and schizophrenia—thereby assessing potential differences in the activation of stigma-related responses across conditions. Finally, we examined the role of individual-level variables, including mental health knowledge and prior contact with people with mental health conditions, as potential factors associated with variations in social distance.

Because generative AI systems have been shown to reproduce, and potentially amplify, social biases embedded in their training data and in the prompts used to guide image generation (Bianchi et al., 2022; Birhane et al., 2021; Luccioni et al., 2023), we reasoned that explicitly incorporating stereotype-related content into prompts should shape the resulting visual representations of mental disorders and, in turn, viewers' stigma-related responses to them.

Based on the existing literature on public stigma, we hypothesized that: images generated under the prompt condition would be associated with higher levels of social distance compared to the no-prompt condition; this effect would be more pronounced for disorders traditionally associated with stronger stigma, such as schizophrenia; higher levels of mental health knowledge would be associated with lower levels of social distance; prior contact with individuals with mental health conditions would be associated with lower levels of social distance.

2. Methods

2.1. Study design

This study employed a within-subject experimental design to examine the impact of AI-generated visual representations of mental disorders on social distance. Participants were exposed to a series of images depicting three mental health conditions—depression, bipolar disorder, and schizophrenia—generated under two conditions: with and without explicit stereotypical prompts. These conditions were selected because they are commonly included within the category of severe mental illness (SMI), which encompasses schizophrenia spectrum disorders, bipolar disorder, and major depressive disorder when associated with significant functional impairment (Ruggeri et al., 2000; Kessler et al., 2005). For each disorder, two images were created under different conditions. In the prompt condition (P), images were generated using prompts that explicitly incorporated widely documented stereotypes associated with the disorder, whereas in the no-prompt condition (NP), images were generated without any stereotypical or descriptive guidance. This resulted in a total of six experimental stimuli (3 disorders $\times$ 2 conditions). Participants evaluated all stimuli, allowing for within-subject comparisons across both disorder type and prompt condition.

The stereotypes used in the prompt condition were derived from the scientific literature on public stigma in mental health and were tailored to each diagnostic category. For schizophrenia, prompts incorporated stereotypes related to dangerousness, unpredictability, social distance, attributions of personal responsibility (e.g., perceived lack of control or moral weakness), and poor prognosis, consistently identified as the most

salient features in public perceptions of this condition (Angermeyer and Matschinger, 2004; Wood et al., 2014; Valery and Prouteau, 2020). For bipolar disorder, prompts reflected stereotypes associated with emotional instability, unpredictability, impaired self-control, and fluctuating functioning. These features are consistent with broader stigma domains such as unpredictability and perceived lack of control, as well as with disorder-specific representations of bipolar disorder described in the literature (Corrigan and Watson, 2002; Ellison et al., 2015; Favre et al., 2023). For depression, prompts were designed to reflect commonly reported stigma-related beliefs, including attributions of personal responsibility (i.e., patient blame), social withdrawal, and perceptions of reduced functioning and competence, as well as the belief that individuals could "pull themselves together," as documented in the literature (Crisp et al., 2000; Wood et al., 2014; Kågström et al., 2025).

The condition-specific stereotypes identified from the literature were translated into structured textual prompts to guide AI-based image generation. For each mental health condition (i.e., depression, bipolar disorder, schizophrenia), prompts were constructed by explicitly incorporating predefined stigma-related attributes (e.g., dangerousness, unpredictability, lack of control, social withdrawal, reduced functioning), operationalized into concise and standardized textual descriptions suitable for input into the image generation system (Table 1). Each prompt was formulated following a consistent structure, including: (a) reference to the diagnostic category; (b) explicit inclusion of one or more stereotype-related attributes. In the no-prompt condition, the AI model was provided exclusively with the diagnostic label (e.g., "schizophrenia," "bipolar disorder," "depression"), without any additional descriptive, contextual, or stereotype-related information, ensuring that no explicit stereotype-related attributes were included in the prompt, beyond the diagnostic label itself. It should be noted that the diagnostic label itself may still implicitly activate stereotypical representations within the model; however, no additional stereotype-related descriptors were provided in this condition.

All images were generated using a large language model (OpenAI, 2024). To ensure transparency and reproducibility, the AI-generated images used as experimental stimuli are available in Supplementary Table 1.

Table 1
Stereotype domains used for AI prompt construction, by diagnostic category.

Disorder	Stereotype domain	Specific content (operationalized for prompts)
Schizophrenia (Crisp et al., 2000; Corrigan and Watson, 2002; Angermeyer and Matschinger, 2004; Angermeyer and Dietrich, 2006; Wood et al., 2014a)	Dangerousness	Perceived as violent, threatening to others, requiring containment or institutionalization
	Unpredictability	Seen as unpredictable, unstable, difficult to understand, or lacking behavioral control
	Reduced competence / Incompetence	Viewed as cognitively impaired, unable to make sound decisions, or incapable of functioning independently
	Poor prognosis	Perceived as chronic and treatment-resistant, with limited prospects for recovery or rehabilitation
Depression (Crisp et al., 2000; Wood et al., 2014a; Kågström et al., 2025)	Responsibility / Blame	Attributions of personal weakness, lack of willpower, or responsibility for the condition; "pull yourself together" beliefs
	Social withdrawal	Perceived as socially withdrawn, difficult to interact with, or isolated
	Reduced competence	Viewed as less productive, less successful, less self-caring, and generally impaired in functioning
	Other attributes	Sometimes associated with laziness, passivity, or introversion
Bipolar disorder (Ellison et al., 2015; Favre et al., 2023)	Unpredictability	Perceived as emotionally unstable, unreliable, and difficult to understand
	Lack of control / Emotional instability	Seen as unable to regulate emotions, excessively volatile, or lacking control over mood and behavior
	Reduced functioning	Viewed as unable to maintain stable functioning, work performance, or everyday responsibilities
	Dangerousness	Occasionally associated with impulsivity and potential risk, although less central than in schizophrenia
	Treatment dependency	Perceived as requiring ongoing medication or treatment to maintain stability
	Creativity (positive stereotype)	Associated with creativity, artistic talent, or heightened intelligence

2.2. Measures

Participants also completed a set of self-report measures assessing sociodemographic characteristics, mental health knowledge, and social distance. Sociodemographic variables included age, gender, educational level, and professional background (where applicable). Additional variables relevant to mental health literacy and stigma—such as prior contact with individuals with mental health conditions and previous training or education in mental health—were also collected.

Social distance was assessed as the primary outcome variable of the study. After viewing each image, participants rated their desired level of social distance toward the depicted individual. Responses were collected using a structured self-report instrument consisting of four items assessing behavioral intentions (i.e., willingness to approach, trust, work with, and engage in conversation with the person depicted). These items reflect commonly used indicators of social distance in stigma research and are conceptually aligned with established measures such as the Reported and Intended Behaviour Scale (Evans-Lacko et al., 2011; Pingani et al., 2016b). Each item was rated on a 5-point Likert scale, and responses were aggregated to compute a composite social distance score, with higher values reflecting greater willingness to engage with the depicted individual, corresponding to lower desired social distance and, therefore, lower levels of stigma. For ease of interpretation, scores were coded so that higher values indicate more positive social responses (i.e., reduced stigma). This composite score was used as the sole outcome measure throughout all analyses. The scale showed excellent internal consistency, Cronbach's $\alpha = .92$ (95% CI [.90, .93]), computed on the composite social distance score aggregated across the six experimental stimuli. Internal consistency was also good to excellent when assessed separately for each of the six stimuli (α ranging from .79 to .93, mean $\alpha = .87$), supporting the reliability of the measure across all experimental conditions.

Mental health knowledge was assessed using the Italian version of the Mental Health Knowledge Schedule (MAKS-I) (Evans-Lacko et al., 2010; Pingani et al., 2019), a widely used instrument designed to evaluate stigma-related knowledge, particularly in relation to help-seeking, recognition, and understanding of mental health conditions. The scale includes items covering domains such as help-seeking, recognition of mental disorders, support, employment, treatment, and recovery. Responses are provided on a Likert-type scale, with higher scores indicating greater mental health knowledge. In line with previous

research, MAKS-I scores were computed by aggregating item responses. The psychometric properties of the Italian version have been previously examined and support its use in both general and clinical populations.

2.3. Statistical analysis

All statistical analyses were conducted using JASP. Descriptive statistics were first computed to characterize the sample and study variables. Continuous variables (e.g., age, MAKS-I scores, social distance scores) were summarized using means and standard deviations (M, SD), while categorical variables (e.g., gender, education level, prior contact with mental health conditions) were described using frequencies and percentages.

Preliminary analyses were performed to assess data quality and distributional assumptions. Normality of continuous variables was evaluated using skewness and kurtosis indices, as well as visual inspection of histograms and Q–Q plots.

The primary analysis examined the effect of experimental condition and diagnostic category on social distance scores using a repeated-measures analysis of variance (RM-ANOVA). Specifically, a 2 × 3 within-subjects design was implemented, with condition (prompt vs. no-prompt) and diagnostic category (depression, bipolar disorder, schizophrenia) as within-subject factors. Mauchly's test was used to assess the assumption of sphericity; when violated, Greenhouse–Geisser corrections were applied. In addition to the primary RM-ANOVA, paired-samples t tests were conducted as follow-up analyses to examine overall differences between conditions and disorder-specific effects.

Main effects of condition and diagnostic category, as well as their interaction, were evaluated. Significant effects were further explored using post hoc pairwise comparisons with Bonferroni correction for multiple testing. Effect sizes were reported using partial eta squared (η^2p) for ANOVA results and Cohen's d for pairwise comparisons.

To explore the potential influence of individual characteristics, secondary analyses were conducted to examine associations between individual-level variables (e.g., mental health knowledge and prior exposure to mental health events) and both social distance scores and condition-related difference scores.

Difference scores were computed as Prompt − No Prompt, with positive values indicating lower social distance (i.e., lower stigma) in the prompt condition, and negative values indicating higher social distance (i.e., greater stigma) in the prompt condition relative to the no-prompt condition. These scores were used in secondary analyses to quantify the magnitude of the prompt effect and to examine associations with individual-level variables.

Associations between difference scores and individual characteristics (e.g., mental health knowledge and prior exposure to mental health events) were examined using Pearson's correlations, linear regression models, and independent-samples t tests, as appropriate. For analyses involving prior exposure to mental health or stigma-related events, only participants who provided a definitive response ("Yes" or "No") were included. Participants who selected "Do not remember" or "Prefer not to answer" were excluded from inferential analyses. Finally, exploratory analyses further examined the relationship between mental health knowledge and social distance scores using Pearson's correlation coefficients. Statistical significance was set at $p < .05$ (two-tailed) for all analyses.

2.4. Sample size

An a priori power analysis was conducted for the primary analysis using a repeated-measures design with two within-subject factors: condition (prompt vs. no-prompt) and diagnostic category (depression, bipolar disorder, schizophrenia). The analysis focused on the interaction between condition and diagnostic category. Assuming a small-to-medium effect size for the interaction (Cohen's $f = 0.15$), an alpha level of .05, statistical power of .80, six repeated measurements, a moderate correlation among repeated measures ($r = .50$), and a conservative nonsphericity correction ($\varepsilon = .75$), the required sample size was estimated to be approximately 170 participants. Under less conservative assumptions (e.g., a medium effect size, $f = 0.20$), the required sample size was estimated to be approximately 100 participants, whereas assuming a larger effect size ($f = 0.25$), the required sample size would decrease to approximately 70 participants.

3. Results

3.1. Sample characteristics

A total of 311 participants completed the study and were included in the analyses. The majority of participants were aged between 18 and 30 years (N = 193; 62.1%), identified as women (N = 221; 71.1%), and reported an upper secondary or university level of education (N = 274; 88.1%). In terms of occupational status, the largest group consisted of students (N = 142; 45.7%). Finally, 51.5% (N = 160) reported previous exposure to mental health or stigma-related events.

Insert Table 2 about here

3.2. Descriptive statistics and difference scores

Descriptive statistics were calculated for all main study variables, including mental health knowledge (MAKS) and social distance scores across conditions and diagnostic categories, along with internal consistency (Cronbach's α) for the social distance scale in each condition (Table 3). The mean MAKS score was 23.13 (SD = ±3.16). The mean difference score for depression was negligible (Δ = 0.02, SD = ±0.60), indicating no meaningful change between conditions. In contrast, bipolar disorder showed a large positive difference (Δ = 1.30, SD = ±1.11), indicating lower social distance in the prompt condition. For schizophrenia, the difference score was negative (Δ = −0.63, SD = ±0.99), indicating higher social distance in the prompt condition.

3.3. H1. effect of prompt condition on social distance

A paired-samples t-test indicated a significant difference between

Table 2
Sample characteristics (N = 311).

Variable	Category	n	%
Age group	18–30 years	193	62.06
	31–50 years	42	13.50
	51 years and older	76	24.44
Gender identity	Men	88	28.30
	Women	221	71.06
	Prefer not to answer / other	2	0.64
Educational level	Lower education	23	7.40
	Upper secondary / university	274	88.10
	Doctoral / specialization	14	4.50
Occupational status	Full-time work	93	29.90
	Part-time work	24	7.72
	Student	142	45.66
	Work and study	25	8.04
	Neither working nor studying	24	7.72
	Prefer not to answer	3	0.96
Participation in mental health or stigma-related events	Yes	160	51.45
	No	138	44.37
	Do not remember	12	3.86
	Prefer not to answer	1	0.32

Table 3
Descriptive statistics for MAKS and condition-specific social distance scores.

Variable	M	SD	Median	Min	Max	Cronbach's α
MAKS total score	23.13	±3.155	23	14	30	-
Depression – Prompt	4.09	±0.90	4.25	1	5	0.79
Depression – No Prompt	4.07	±0.86	4.25	1	5	0.82
Δ Depression	0.02	±0.60	0	−2	2.25	-
Bipolar – Prompt	3.46	±1.10	3.5	1	5	0.91
Bipolar – No Prompt	2.16	±1	2	1	5	0.88
Δ Bipolar	1.30	±1.11	1.25	−3	4	-
Schizophrenia – Prompt	2.99	±1.11	3	1	5	0.90
Schizophrenia – No Prompt	3.62	±1.13	3.75	1	5	0.93
Δ Schizophrenia	−0.63	±0.99	−0.5	−4	2	-
Mean social distance score across stimuli	3.40	±0.78	3.38	1.33	5	0.92
Prompt overall	3.51	±0.85	3.5	1	5	-
No Prompt overall	3.28	±0.80	3.25	1.08	5	-

Note: "—" indicates rows for which a 4-item Cronbach's α is not directly applicable (MAKS is a separate instrument; Δ and "overall" rows are difference/aggregate scores across multiple conditions).

conditions, $t(310) = 7.75$, $p < .001$. Participants reported higher scores in the prompt condition ($M = 3.51$, $SD = \pm 0.85$) compared to the no-prompt condition ($M = 3.28$, $SD = \pm 0.80$), with a mean difference of 0.23 (95% CI [0.17, 0.29], $d = 0.44$). Given that higher scores reflect lower desired social distance, this pattern suggests that images generated using stereotype-based prompts were, on average, associated with slightly lower levels of stigma. This finding runs counter to the initial hypothesis, which predicted higher levels of social distance (i.e., greater stigma) in the prompt condition.

H2. Disorder-specific effects

To further examine this unexpected overall effect, analyses were conducted separately for each diagnostic category (Table 4), and different patterns emerged across conditions. For depression, no significant difference was observed between the prompt and no-prompt conditions, $t(310) = 0.64$, $p = .521$, $d = 0.04$. For bipolar disorder, a statistically significant difference was found, $t(310) = 20.75$, $p < .001$, $d = 1.18$, with higher scores in the prompt condition compared to the no-prompt condition. For schizophrenia, a statistically significant difference in the opposite direction was observed, $t(310) = -11.27$, $p < .001$, $d = -0.64$, with lower scores in the prompt condition, indicating higher social distance.

3.4. H3. condition × diagnosis interaction

A 2 × 3 repeated-measures ANOVA was conducted to examine the effects of condition (prompt vs. no-prompt) and diagnosis (depression, bipolar disorder, schizophrenia) on social distance scores. The analysis revealed a significant main effect of condition, $F(1, 310) = 60.01$, $p < .001$, $\eta^2p = .162$, indicating overall differences between prompt and no-prompt presentations. A significant main effect of diagnosis also emerged, $F(2, 620) = 388.84$, $p < .001$, $\eta^2p = .556$, reflecting variability in social distance across the three disorders. A significant condition × diagnosis interaction was observed, $F(2, 620) = 348.24$, $p < .001$, $\eta^2p = .529$, indicating that the effect of prompt condition varied across diagnostic categories. This interaction indicates that the effect of prompt condition is not uniform across diagnoses.

3.5. H4. mental health knowledge

Mental health knowledge was positively associated with social distance scores, $r = .19$, $p < .001$, indicating that higher levels of knowledge were related to lower levels of stigma. Further analyses using linear regression models were conducted to examine whether mental health knowledge predicted the magnitude of the prompt effect for each diagnostic category. Mental health knowledge did not significantly predict the difference scores for depression or bipolar disorder (ps > .15). In contrast, a significant association was observed for schizophrenia, $B = -0.049$, $SE = 0.018$, $p = .005$, 95% CI [−0.084, −0.015], $R^2 = .025$. The negative coefficient indicates that higher levels of mental health knowledge were associated with more negative difference scores (Δ = Prompt − No Prompt), corresponding to relatively higher social distance (i.e., greater stigma) in the prompt condition compared to the no-prompt condition.

3.6. H5. exposure to mental health events

Participants who reported prior exposure to mental health or stigma-related events (limited to those providing “Yes” or “No” responses) showed significantly higher social distance scores compared to those without such exposure, $t(296) = 3.08$, $p = .002$, $d = 0.36$. Given that higher scores indicate lower desired social distance, this pattern corresponds to lower levels of stigma among participants with previous exposure. Additional analyses were conducted to examine whether prior exposure to mental health or stigma-related events was associated with the magnitude of the prompt effect for each diagnostic category. Independent-samples t tests comparing difference scores (Δ = Prompt − No Prompt) between participants with and without prior exposure did not reveal significant differences for depression ($p = .084$), bipolar disorder ($p = .254$), or schizophrenia ($p = .332$).

4. Discussion

The present study aimed to examine whether AI-generated images of mental disorders influence social distance depending on the use of stereotype-based prompts, whether this effect varies across diagnostic categories (depression, bipolar disorder, schizophrenia), and whether individual-level factors—namely mental health knowledge and prior exposure to mental health-related experiences—are associated with these responses.

4.1. H1. effect of stereotype-based prompts on social distance

Contrary to our primary hypothesis (H1), results showed that images generated using stereotype-based prompts were associated with lower levels of desired social distance compared to images generated without prompts. Given that higher scores reflect more positive social responses, this indicates that stereotype-based prompts were, on average, associated with slightly reduced stigma.

This finding diverges from well-established models of public stigma,

Table 4
Disorder-specific comparisons between prompt and no-prompt conditions.

Disorder	Prompt M (±SD)	No Prompt M (±SD)	Mean difference	t (df)	p	Cohen's d
Depression	4.09 (±0.90)	4.07 (0.86)	0.02	0.64 (310)	.521	0.04
Bipolar disorder	3.46 (±1.10)	2.16 (1.00)	1.30	20.75 (310)	< .001	1.18
Schizophrenia	2.99 (±1.11)	3.62 (1.13)	−0.63	−11.27 (310)	< .001	−0.64

which posit that the activation of stereotypical beliefs—particularly those related to dangerousness, unpredictability, and personal responsibility—should increase avoidance and social distancing (Corrigan and Watson, 2002; Angermeyer and Matschinger, 2004). In contrast, the present results suggest that the translation of stereotypes into AI-generated visual content does not necessarily produce a straightforward amplification of stigma, but may instead reflect more complex and context-dependent pathways linking stereotypes to discriminatory responses (Kågström et al., 2025). A parsimonious interpretation, grounded in the present data, is that prompt-based images may have resulted in more structured or interpretable representations, potentially reducing ambiguity compared to the no-prompt condition, where the diagnostic label alone may have activated more diffuse or internally generated representations. However, given that the overall effect was small ($d = 0.44$) and substantially qualified by the interaction with diagnosis, this finding should not be interpreted as evidence of a general destigmatizing effect of stereotype-based prompts.

4.2. H2. disorder-specific effects

Findings related to H2 indicate that the effect of stereotype-based prompts varied substantially across diagnostic categories, and therefore the hypothesis was only partially supported.

For depression, no significant differences were observed between conditions. This is consistent with prior literature suggesting that depression is generally perceived as less threatening and more socially acceptable compared to other mental disorders (Kågström et al., 2025; Wood et al., 2014). The stereotype domains associated with depression—such as responsibility or reduced functioning—may not translate into visually salient cues capable of modifying immediate behavioral intentions in this type of task.

For bipolar disorder, results showed a large effect in the opposite direction to that hypothesized, with significantly lower social distance in the prompt condition. This finding contrasts with previous evidence emphasizing unpredictability and instability as central components of bipolar disorder stigma (Ellison et al., 2015; Favre et al., 2023). Importantly, in the present study, prompts for bipolar disorder included not only negative attributes but also positive or ambivalent stereotypes, such as creativity and heightened intelligence. The inclusion of these elements may have contributed to more favorable representations, consistent with literature suggesting that mixed stereotypes can attenuate stigma under certain conditions. However, because each condition was represented by a single image, this interpretation remains speculative: uncontrolled visual features of the specific stimulus used (e.g., sex of the depicted individual, apparent emotional intensity, extent of the person shown, lighting) cannot be ruled out as alternative explanations for this finding, and it should be treated as one plausible account among others rather than an established explanation.

For schizophrenia, results were consistent with theoretical expectations. Images generated using stereotype-based prompts were associated with significantly higher social distance, indicating increased stigma. The prompts explicitly incorporated highly salient stigma-related domains—particularly dangerousness, unpredictability, and poor prognosis—which are well-documented drivers of public stigma toward schizophrenia (Angermeyer and Matschinger, 2004; Valery and Prouteau, 2020).

Taken together, these findings suggest that the impact of stereotype-based prompts is not determined solely by diagnostic category, but also by the content and valence of the stereotypes included. While H2 is supported for schizophrenia at a directional level, it is clearly contradicted by the pattern observed for bipolar disorder, where the largest effect emerged in the opposite direction.

4.3. H3. condition × diagnosis interaction

The significant condition × diagnosis interaction represents a central finding of the study and provides an integrated framework for interpreting H1 and H2. The interaction indicates that the effect of stereotype-based prompts is not uniform across mental disorders but varies in both magnitude and direction. This finding is consistent with multidimensional models of stigma, which emphasize that different disorders are associated with distinct stereotype profiles and emotional responses (Corrigan and Watson, 2002). In the present study, schizophrenia was associated with a pattern consistent with fear-based stigma (increased social distance under stereotype activation), bipolar disorder with a pattern suggestive of ambivalent or mixed stereotyping (reduced social distance under prompt condition), and depression with a relatively neutral pattern. These differentiated effects underscore the importance of moving beyond aggregate analyses and considering diagnosis-specific mechanisms when examining stigma in visual and AI-mediated contexts. At the same time, it may be useful to consider a complementary, symptom-oriented perspective. Although diagnostic categories are associated with characteristic symptom profiles, stigma-related responses may be driven more directly by the specific features represented (e.g., negative symptoms, psychotic experiences, or affective instability) rather than by the diagnostic label per se. From this perspective, future research could explore how different symptom-level representations—particularly those that are less stereotypical or more ambiguous—shape stigma-related responses across and within diagnostic categories.

4.4. H4. mental health knowledge

Consistent with H3 (as originally formulated), mental health knowledge was associated with lower levels of stigma overall, as indicated by the positive correlation with social distance scores. This aligns with previous research showing that greater knowledge is linked to more favorable attitudes toward individuals with mental health conditions (Evans-Lacko et al., 2010; Pingani et al., 2019). However, knowledge did not uniformly moderate the effect of stereotype-based prompts. A significant association was observed only for schizophrenia, where higher knowledge predicted more negative difference scores—that is, relatively higher social distance in the prompt condition. This finding suggests that the role of knowledge may differ depending on the context and the type of stimulus. In the case of schizophrenia, higher knowledge may be associated with greater sensitivity to stereotype-consistent cues or with more consolidated representations of the disorder, which may amplify responses to stereotypical depictions. Importantly, this result indicates that knowledge alone may not be sufficient to buffer against stigma in contexts where highly salient stereotypes are activated.

4.5. H5. exposure to mental health events

Participants with prior exposure to mental health or stigma-related events reported lower levels of stigma overall, as reflected in higher social distance scores. This finding is in line with a substantial body of literature demonstrating that contact is associated with reduced stigma (Corrigan and Watson, 2002). However, prior exposure did not significantly moderate the effect of prompt condition for any of the diagnostic categories. This suggests that while exposure may influence baseline attitudes, it does not necessarily alter responses to different types of AI-generated visual representations.

This pattern indicates that the mechanisms underlying general stigma reduction (e.g., through contact) and those underlying responses to specific visual stimuli may be partially distinct. In particular, exposure may be more strongly related to changes in internalized or self-stigma processes and general attitudes, rather than to immediate reactions to externally presented representations. This distinction may help explain why prior exposure was associated with overall lower stigma but did not buffer against the effects of stereotype-based visual cues.

4.6. Implications

The present findings have several implications for research and practice in the field of stigma and mental health (Del Rosal et al., 2021). First, they challenge the assumption that stereotype-based representations will uniformly increase stigma. Instead, the effects appear to depend on both the diagnostic category and the specific content of the stereotypes included. This is consistent with recent work highlighting that stigma is shaped by complex and context-dependent representational processes rather than by the mere activation of stereotypes (Link and Phelan, 2001; Fiske et al., 2002; Angermeyer and Dietrich, 2006).

Second, they highlight the potential for AI-generated imagery to produce heterogeneous effects, including both stigma-enhancing and stigma-reducing outcomes. This is particularly relevant given the increasing use of generative AI in communication and media contexts, where algorithmic systems have been shown to reproduce, amplify, or transform social biases depending on training data and prompt design (Bianchi et al., 2022; Luccioni et al., 2023). Taken together, these findings suggest that generative AI functions less as an independent source of stigma than as a mirror and potential amplifier of existing societal stereotypes, reproducing culturally embedded representations of mental illness that are already present in its training data and in the assumptions encoded in user prompts. This reframing situates the present findings within the broader literature on algorithmic bias and extends their relevance beyond image-generation technology itself, to the social representations that such technology reflects.

Third, the findings underscore the importance of carefully designing prompts and monitoring AI-generated outputs, especially for highly stigmatized conditions such as schizophrenia, where stereotype activation appears to have a consistently negative effect. This aligns with emerging discussions on the ethical governance of generative AI and the need for bias-aware design and evaluation frameworks (Angermeyer and Matschinger, 2003; Raji et al., 2020; Weidinger et al., 2022). These findings also carry practical implications for the various stakeholders who increasingly rely on AI-generated imagery in mental health communication. Researchers should consider systematically evaluating the visual content of AI-generated stimuli before using them in stigma-related research. Clinicians and advocacy organizations producing educational or public-facing materials should be aware that even minimally specified prompts may still activate diagnosis-specific stereotypes, and that prompt design is not a neutral technical step but a communicative choice with potential stigma-related consequences. Developers of generative AI tools and platforms, in turn, may consider incorporating bias-awareness safeguards or guidance for mental-health-related image generation, given the documented tendency of these systems to reproduce embedded societal stereotypes.

4.7. Limitations and future directions

Several limitations should be acknowledged when interpreting the findings of the present study.

The use of self-report measures of social distance, although consistent with established approaches in stigma research, may be subject to social desirability bias. Participants may have responded in a manner perceived as socially acceptable rather than fully reflecting their spontaneous attitudes, particularly given the increasing public awareness around mental health stigma. While the within-subject design partially mitigates this concern by focusing on relative differences across conditions, it remains possible that the absolute levels of social distance were underestimated. Future studies could complement self-report data with implicit or behavioral measures, such as reaction time paradigms, choice-based tasks, or ecological simulations, to capture more automatic or less consciously regulated responses.

Second, the characteristics of the sample may limit the generalizability of the findings. The sample was predominantly young, female, and highly educated, with a substantial proportion of students. These characteristics are typically associated with lower levels of stigma and higher mental health literacy, which may have influenced both baseline responses and sensitivity to experimental manipulations. In addition, the relatively high proportion of participants reporting prior exposure to mental health-related experiences may have further contributed to reduced overall stigma. Future research should aim to replicate these findings in more diverse and representative samples, including populations with different age distributions, educational levels, cultural backgrounds, and levels of familiarity with mental health.

The study focused on a limited set of diagnostic categories—depression, bipolar disorder, and schizophrenia—selected due to their relevance within the broader category of severe mental illness. However, stigma varies substantially across mental health conditions, and different patterns may emerge for other diagnoses, such as anxiety disorders, personality disorders, or substance use disorders. In particular, conditions associated with moral judgment or perceived controllability may show different responses to stereotype-based visual representations. Expanding the range of diagnostic categories would allow for a more comprehensive understanding of how AI-generated imagery interacts with disorder-specific stigma profiles.

The operationalization of stereotype-based prompts represents an important methodological constraint. Although prompts were systematically derived from the literature and structured to ensure consistency across conditions, they necessarily involved a selection and simplification of complex stigma domains. Moreover, the inclusion of both negative and positive or ambivalent stereotypes—particularly in the case of bipolar disorder—may have influenced the direction and magnitude of the observed effects. Future studies should more systematically manipulate the valence, intensity, and combination of stereotype components within prompts, in order to disentangle their specific contributions to stigma-related responses.

Relatedly, each condition in the present study was represented by a single image. As a result, potential confounding features of the specific stimuli used—such as the sex of the depicted individual, apparent level of sadness, extent of the person shown, and lighting—were not controlled and cannot be ruled out as alternative explanations for the observed disorder-specific effects. This concern applies to all three diagnostic categories and may be particularly relevant to the bipolar disorder findings, where the largest and most unexpected effect emerged. Relatedly, the present findings are also based on a relatively small set of images produced with a single generative AI system and a specific set of prompts. As such, they cannot be assumed to generalize across different prompt formulations, image-generation models, or future AI systems, which may embed different biases and produce qualitatively different visual representations. The study did not directly examine the visual characteristics of the AI-generated images, which represent a key intermediate mechanism between prompt content and participant responses. While the experimental manipulation was defined at the level of prompts, the resulting images may have varied along multiple dimensions—such as emotional expression, realism, contextual cues, or perceived threat—that were not formally assessed. Future research should incorporate image-level analyses, including both objective features (e.g., facial expressions, color composition, contextual elements) and subjective evaluations (e.g., perceived realism, emotional valence, perceived dangerousness), to better understand how specific visual properties mediate the relationship between prompt design and stigma outcomes.

Importantly, future studies should employ several images per condition, allowing stimulus-specific variance to be statistically disentangled from diagnosis-specific effects.

Sixth, the study design does not allow for disentangling whether the observed effects are driven primarily by the content of the prompts, the characteristics of the generated images, or the interaction between the two. In particular, the no-prompt condition—based solely on diagnostic labels—may have activated participants' pre-existing mental representations, which could differ substantially across individuals. As such, the

comparison between conditions reflects not only differences in AI-generated content but also differences in the cognitive processes engaged by participants. Future studies could address this limitation by including additional control conditions (e.g., standardized human-generated images, neutral prompts, or explicitly anti-stigma prompts) to better isolate the contribution of each component.

The cross-sectional and experimental nature of the study captures immediate responses to visual stimuli but does not provide information on the persistence or real-world impact of these effects. It remains unclear whether the observed differences in social distance would translate into longer-term attitudinal changes or actual behavioral outcomes. Longitudinal designs and field-based studies are needed to assess the durability and ecological validity of these effects, particularly in real-world contexts such as social media, education, and public communication.

Building on these limitations, several directions for future research can be identified.

A first priority is to investigate the mechanisms underlying the observed heterogeneity across diagnostic categories. In particular, future studies should examine how different types of stereotypes—negative, positive, and ambivalent—interact with visual representations to shape stigma-related responses. This could involve experimentally manipulating specific stereotype domains (e.g., dangerousness vs. creativity) within prompts and assessing their independent and combined effects. A second important direction concerns the role of prompt design as a modifiable factor. The present findings suggest that prompts do not merely transmit stereotypes but actively shape the nature of visual representations and their social impact. Future research should explore the potential of recovery-oriented, strength-based, or anti-stigma prompts to generate images that promote more accurate and less stigmatizing representations of mental health conditions. Such work could contribute to the development of evidence-based guidelines for the ethical use of generative AI in mental health communication. A third line of research should focus on individual differences in responses to AI-generated imagery. While the present study examined mental health knowledge and prior exposure, additional factors—such as personal experience with mental illness, professional background, cultural context, and implicit attitudes—may further moderate responses. Understanding these moderating variables could help identify groups that are more vulnerable to stigma-enhancing representations or more responsive to anti-stigma interventions. A fourth direction involves extending the investigation to more ecologically valid settings. Given the widespread use of AI-generated images in digital environments, it is important to examine how such representations are encountered and interpreted in real-world contexts, such as social media platforms, news media, and educational materials. Studies incorporating more naturalistic exposure paradigms or field experiments would provide valuable insights into the practical implications of these findings. Finally, future research should engage more directly with the ethical and societal implications of AI-generated representations of mental health. The present findings highlight both the risks and the potential opportunities associated with these technologies. On the one hand, stereotype-based prompts may reinforce stigma, particularly for highly stigmatized conditions such as schizophrenia. On the other hand, carefully designed prompts may contribute to more nuanced or even destigmatizing representations. Developing frameworks for the responsible design, evaluation, and deployment of AI-generated content in mental health contexts represents a critical area for interdisciplinary research.

Ethics statement

This study was conducted in accordance with internationally recognized ethical standards for research involving human participants, including the principles of the Declaration of Helsinki. The research involved the collection of participants' opinions on visual materials and did not include clinical assessments or diagnostic procedures. According to applicable regulations for minimal-risk, non-clinical research, formal ethical committee approval was not required. Participants were informed about the aims and procedures of the study and provided informed consent prior to participation. Data were collected anonymously and handled in compliance with applicable data protection and confidentiality regulations, including the General Data Protection Regulation (EU 2016/679). The authors confirm that data are available for verification purposes upon reasonable request, while ensuring the protection of participants' confidentiality.

CRediT authorship contribution statement

Luca Pingani: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Nicole Lanzi:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Erika De Marco:** Writing – review & editing, Methodology, Investigation. **Giulia Ferrazzi:** Writing – review & editing, Investigation. **Anna Ghirardello:** Writing – review & editing, Investigation. **Mattia Marchi:** Writing – review & editing, Methodology, Data curation. **Annamaria Nasi:** Writing – review & editing, Supervision. **Giorgia Polizzotto:** Writing – review & editing, Investigation, Conceptualization. **Giulia Pollice:** Writing – review & editing, Investigation. **Maria Chiara Trapasso Colli:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Gian Maria Galeazzi:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors whose names are listed immediately above certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. No funding was received from generative AI companies or related industries for the conduction of this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2026.117450.

Data availability

The data, materials, and analysis code that support the findings of this study are available from the corresponding author upon reasonable request. Access may be granted for research purposes, subject to appropriate justification and in compliance with ethical and data protection requirements.

References

Andalibi, N., Pyle, C., Barta, K., Xian, L., Jacobs, A.Z., Ackerman, M.S., 2023. Conceptualizing algorithmic stigmatization. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Presented at the CHI '23: CHI Conference on Human Factors in Computing Systems. Hamburg Germany. ACM, pp. 1–18. https://doi.org/10.1145/3544548.3580970.

Angermeyer, M.C., Dietrich, S., 2006. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatr. Scand. 113, 163–179. https://doi.org/10.1111/j.1600-0447.2005.00699.x.

Angermeyer, M.C., Matschinger, H., 2004. The stereotype of schizophrenia and its impact on discrimination against people with schizophrenia: results from a

representative survey in Germany. Schizophr. Bull. 30, 1049–1061. https://doi.org/10.1093/oxfordjournals.schbul.a007120.
Angermeyer, M.C., Matschinger, H., 2003. The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder. Acta Psychiatr. Scand. 108, 304–309. https://doi.org/10.1034/j.1600-0447.2003.00150.x.
Betton, V., Borschmann, R., Docherty, M., Coleman, S., Brown, M., Henderson, C., 2015. The role of social media in reducing stigma and discrimination. Br. J. Psychiatry 206, 443–444. https://doi.org/10.1192/bjp.bp.114.152835.
Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., Caliskan, A., 2022. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. https://doi.org/10.48550/ARXIV.2211.03759.
Birhane, A., Prabhu, V.U., Kahembwe, E., 2021. Multimodal datasets: misogyny, pornography, and malignant stereotypes. https://doi.org/10.48550/ARXIV.2110.01963.
Brown, A., Rice, S.M., Rickwood, D.J., Parker, A.G., 2016. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia-Pac. Psychiatry 8, 3–22. https://doi.org/10.1111/appy.12199.
Chan, G., Yanos, P.T., 2018. Media depictions and the priming of mental illness stigma. Stig. Health 3, 253–264. https://doi.org/10.1037/sah0000095.
Corrigan, P.W., Morris, S.B., Michaels, P.J., Rafacz, J.D., Rüsch, N., 2012. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. PS (Wash. DC) 63, 963–973. https://doi.org/10.1176/appi.ps.201100529.
Corrigan, P.W., Watson, A.C., 2002. Understanding the impact of stigma on people with mental illness. World Psychiatry 1, 16–20.
Crisp, A.H., Gelder, M.G., Rix, S., Meltzer, H.I., Rowlands, O.J., 2000. Stigmatisation of people with mental illnesses. Br. J. Psychiatry 177, 4–7. https://doi.org/10.1192/bjp.177.1.4.
Del Rosal, E., González-Sanguino, C., Bestea, S., Boyd, J., Muñoz, M., 2021. Correlates and consequences of internalized stigma assessed through the Internalized Stigma of Mental Illness Scale for people living with mental illness: a scoping review and meta-analysis from 2010. Stig. Health 6, 324–334. https://doi.org/10.1037/sah0000267.
Ellison, N., Mason, O., Scior, K., 2015. Public beliefs about and attitudes towards bipolar disorder: testing theory based models of stigma. J. Affect. Disord. 175, 116–123. https://doi.org/10.1016/j.jad.2014.12.047.
Evans-Lacko, S., Little, K., Meltzer, H., Rose, D., Rhydderch, D., Henderson, C., Thornicroft, G., 2010. Development and psychometric properties of the Mental health knowledge schedule. Can. J. Psychiatry 55, 440–448. https://doi.org/10.1177/070674371005500707.
Evans-Lacko, S., Rose, D., Little, K., Flach, C., Rhydderch, D., Henderson, C., Thornicroft, G., 2011. Development and psychometric properties of the reported and Intended Behaviour Scale (RIBS): a stigma-related behaviour measure. Epidemiol. Psychiatr. Sci. 20, 263–271. https://doi.org/10.1017/S2045796011000308.
Favre, S., Aubry, J.-M., Richard-Lepouriel, H., 2023. Perceived public stigma and perceived public exposure by persons living with bipolar disorder: a qualitative study. Int. J. Soc. Psychiatry 69, 378–387. https://doi.org/10.1177/00207640221093495.
Fiske, S.T., Cuddy, A.J.C., Glick, P., Xu, J., 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. J. Personal. Soc. Psychol. 82, 878–902. https://doi.org/10.1037/0022-3514.82.6.878.
Fox, A.B., Earnshaw, V.A., Taverna, E.C., Vogt, D., 2018. Conceptualizing and measuring mental illness stigma: the mental illness stigma framework and critical review of measures. Stig. Health 3, 348–376. https://doi.org/10.1037/sah0000104.
Kågström, A., Guerrero, Z., Aliev, A.A., Tomášková, H., Rüsch, N., Ouali, U., Thornicroft, G., Sartorius, N., Winkler, P., 2025. Mental health stigma and its consequences: a systematic scoping review of pathways to discrimination and adverse outcomes. EClinicalMedicine 89, 103588. https://doi.org/10.1016/j.eclinm.2025.103588.
Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., Walters, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National comorbidity Survey replication. Arch. Gen. Psychiatry 62, 593. https://doi.org/10.1001/archpsyc.62.6.593.
Knaak, S., Mantler, E., Szeto, A., 2017. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. Heal. Manage Forum 30, 111–116. https://doi.org/10.1177/0840470416679413.
Link, B.G., Phelan, J.C., 2001. Conceptualizing Stigma. Annu Rev. Sociol. 27, 363–385. https://doi.org/10.1146/annurev.soc.27.1.363.
Livingston, J.D., Boyd, J.E., 2010. Correlates and consequences of internalized stigma for people living with mental illness: a systematic review and meta-analysis. Soc. Sci. Med. 71, 2150–2161. https://doi.org/10.1016/j.socscimed.2010.09.030.
Luccioni, A.S., Akiki, C., Mitchell, M., Jernite, Y., 2023. Stable bias: analyzing societal representations in diffusion models. https://doi.org/10.48550/ARXIV.2303.11408.
Morgan, A.J., Reavley, N.J., Ross, A., Too, L.S., Jorm, A.F., 2018. Interventions to reduce stigma towards people with severe mental illness: systematic review and meta-analysis. J. Psychiatr. Res. 103, 120–133. https://doi.org/10.1016/j.jpsychires.2018.05.017.
OpenAI. ChatGPT (GPT-4 version) [Large language model]. https://chat.openai.com.
Pingani, L., Catellani, S., Del Vecchio, V., Sampogna, G., Ellefson, S.E., Rigatelli, M., Fiorillo, A., Evans-Lacko, S., Corrigan, P.W., 2016a. Stigma in the context of schools: analysis of the phenomenon of stigma in a population of university students. BMC Psychiatry 16, 29. https://doi.org/10.1186/s12888-016-0734-8.
Pingani, L., Evans-Lacko, S., Luciano, M., Del Vecchio, V., Ferrari, S., Sampogna, G., Croci, I., Del Fatto, T., Rigatelli, M., Fiorillo, A., 2016b. Psychometric validation of the Italian version of the Reported and Intended Behaviour Scale (RIBS). Epidemiol. Psychiatr. Sci. 25, 485–492. https://doi.org/10.1017/S2045796015000633.
Pingani, L., Sampogna, G., Evans-Lacko, S., Gozzi, B., Giallonardo, V., Luciano, M., Galeazzi, G.M., Fiorillo, A., 2019. How to measure knowledge about mental disorders? Validation of the Italian version of the MAKS. Community Ment. Health J. 55, 1354–1361. https://doi.org/10.1007/s10597-019-00416-6.
Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P., 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. https://doi.org/10.48550/ARXIV.2001.00973.
Ruggeri, M., Leese, M., Thornicroft, G., Bisoffi, G., Tansella, M., 2000. Definition and prevalence of severe and persistent mental illness. Br. J. Psychiatry 177, 149–155. https://doi.org/10.1192/bjp.177.2.149.
Stuart, H., 2006. Media portrayal of mental illness and its treatments: what effect does it have on people with mental illness? CNS Drugs 20, 99–106. https://doi.org/10.2165/00023210-200620020-00002.
Valery, K.-M., Prouteau, A., 2020. Schizophrenia stigma in mental health professionals and associated factors: a systematic review. Psychiatry Res. 290, 113068. https://doi.org/10.1016/j.psychres.2020.113068.
Wahl, O.F., 2003. Media Madness: Public Images of Mental Illness, Second paperback Printing. Rutgers University Press, New Brunswick, New Jersey.
Wang, Z., Doyle, D.M., Zhang, Z., Mohsin, F.Z., Barreto, M., 2026. Structural stigma: a systematic review of existing measures. Stig. Health 11, 141–168. https://doi.org/10.1037/sah0000671.
Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L.A., Rimell, L., Isaac, W., Haas, J., Legassick, S., Irving, G., Gabriel, I., 2022. Taxonomy of risks posed by language models. In: 2022 ACM Conference on Fairness Accountability and Transparency. Presented at the FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency. Seoul Republic of Korea. ACM, pp. 214–229. https://doi.org/10.1145/3531146.3533088.
Wood, L., Birtel, M., Alsawy, S., Pyle, M., Morrison, A., 2014a. Public perceptions of stigma towards people with schizophrenia, depression, and anxiety. Psychiatry Res. 220, 604–608. https://doi.org/10.1016/j.psychres.2014.07.012.