

Article

Not peer-reviewed version

Exploring the Link Between the Emotional Recall Task and Mental Health in Humans and LLMs

[Alessandra Carini](#) , [Enrique Taietta](#) , [Massimo Stella](#) *

Posted Date: 14 October 2025

doi: 10.20944/preprints202510.1021.v1

Keywords: artificial intelligence; cognitive networks; network science; affect; mental health



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Exploring the Link Between the Emotional Recall Task and Mental Health in Humans and LLMs

Alessandra Carini, Enrique Taietta and Massimo Stella * 

CogNosco Lab, Corso Bettini 31, 38068 Rovereto, Trento, Italy

* Correspondence: massimo.stella-1@unitn.it

Abstract

The ability of Large Language Models to recall human emotions provides a novel opportunity to investigate links between memory, affect, and mental health. This study explores whether the Emotional Recall Task (ERT), a free word-association paradigm, can reveal cognitive markers of distress in both humans and Large Language Models (LLMs). Using spreading activation simulations grounded in cognitive network science, we examined how the recall of emotional concepts (e.g., stress, anxiety, depression) relates to psychometric measures of well-being and personality. In Study 1, correlations were tested between activation dynamics and clinical scales (DASS-21, PANAS, Life Satisfaction) in human participants (N=1200) and artificial participants generated by GPT-4, Claude Haiku, and Anthropic Opus. Humans with higher distress scores exhibited stronger, faster, and more persistent activation of negative concepts, supporting theories of rumination and memory bias. GPT-4 approximated human-like trajectories most closely, though with reduced variability. Study 2 linked recall dynamics to Big Five traits, confirming that neuroticism predicted greater activation of negative concepts, while extraversion acted as a protective factor. While LLMs lack autobiographical memory, their semantic activations partially mirrored human associations. These findings demonstrate that network-based spreading activation analysis can reveal cognitive signatures of distress, highlighting also the limits of LLMs in modeling human affect.

Keywords: artificial intelligence; cognitive networks; network science; affect; mental health

1. Introduction

The ability of artificial systems to replicate human-like emotional recall provides a new perspective to evaluate how closely a Large Language Model (LLM) can approximate complex psychological processes [1]. Examples include the logical reasoning behind providing an argument [2], the activation of emotion-laden words in clinical settings [1], or the agreement of naming conventions in social systems [3]. Humans and LLMs might differ significantly in how emotions and cognition interact with each other. In humans, emotions strongly shape cognitive processes, influencing how people perceive the world, retrieve memories, and approach everyday tasks in general [4]. When negative emotional states intensify, leading to distress, anxiety or depressive disorders, human behavior and decision making can be further compromised. Anxiety, for instance, is known to negatively impact executive functioning [5]; depression is closely intertwined with the cognitive structure that sustains it [6] and stress interferes with processes such as attention, memory formation, and recall [7]. These connections highlight how close bonds between emotions and cognition are, especially regarding memory. Cognitive network science offers an effective way to examine these connections, enabling researchers to map the network of associations that support or amplify negative emotional states ([8–10]).

Cognitive network science conceptualizes knowledge as a network of interconnected concepts [11–13], where associations reflect how information is stored and retrieved in a mental lexicon of knowledge [9]. Within this framework, spreading activation models describe how the retrieval of one

concept can trigger cascades of activity across related nodes, allowing emotionally charged words to become more accessible depending on their position and connectivity in the network [8–10]. For example, highly central nodes such as "stress" or "anxiety" may receive disproportionate activation, amplifying their salience and sustaining negative emotional states through recurrent loops of activation. This mechanism provides a quantitative account of memory biases observed in psychological distress, where minor triggers can rapidly propagate toward clusters of negatively valenced concepts. By applying these tools, cognitive network science [11] enables researchers to capture how structural and dynamical properties of semantic networks mirror individual differences in affect, personality, and vulnerability to mental disorders. This one is the approach we follow in this work.

By integrating network metrics of concepts in a network of memory recalls between concepts, this work aims to understand how closely the activation of nodes related to mental distress (such as "anxiety", "stress", and "depression") correlates with measurable psychological indicators of well-being, both at clinical and personality level. Applying network science and spreading activation [10], we focus on the interpretation of network-driven results in the context of mental health. Additionally, we also aim to highlight the difference between humans, whose cognitive representations are shaped by personal history and socio-cognitive factors [13,14], and LLMs, which rely on linguistic training data without direct experience or autobiographical context of events and concepts [15].

2. Literature Review

Cognitive network science are representational models of cognition. Cognitive network science models human knowledge as a network of concepts linked by learned associations [11], enabling formal study of how structure shapes retrieval and reasoning [9,10,13]. Foundational spreading-activation theories [8,16] propose that cueing one concept propagates activation along associative links, increasing the accessibility of nearby nodes; these ideas are now operationalized in computational simulations that estimate how activation unfolds over time across a lexical network. Normative free-association resources such as the Small World of Words (SWOW) [17] provide the backbone topology for such models, while affective lexica with Valence–Arousal–Dominance (VAD) norms [18] annotate nodes with emotional properties, allowing joint analyses of semantics and affect. Within this framework, classic network measures (e.g., degree, shortest paths, clustering, centrality) predict which concepts act as hubs or bridges for activation flow, and hence which ideas are most likely to be retrieved under minimal input [9,10]. When negatively valenced clusters (e.g., anxiety, stress, depression) are densely interconnected, even mild cues can traverse short associative paths and disproportionately energize these hubs, sustaining recurrent activation loops consistent with rumination and other maladaptive dynamics observed in psychopathology.

Personality Traits are linked with mental health. The Big Five Personality Traits, Neuroticism, Conscientiousness, Extraversion, Agreeableness, and Openness [19] have long been linked to psychological well-being and increased risk for psychopathology. Among these, the neuroticism trait is most consistently associated with greater vulnerability to negative emotions and with an elevated risk of developing anxiety and depressive disorders [14,20]. Neurotic individuals tend to simulate past and future problems, therefore experiencing negative mood states in the apparent absence of a plausible cause [21]. From a network science perspective, we can assume that these individuals possess cognitive networks where concepts linked with negative mood states (e.g. "anxiety" or "depression") also have an increased connectivity. This network aspect could lead to quicker or more persistent activation of anxious or depressive thoughts, characteristic of rumination. Rumination is defined as the repetitive and excessive focus on negative emotions, thoughts or events [22]. It has been strongly linked to the development and maintenance of affective disorders such as depression and anxiety [22], and commonly occurs in other mental disorders. In the context of spreading activation models [16], rumination can be interpreted as a self-perpetuating activation loop: once a negative concept is triggered (e.g. concepts/nodes like *depression*, *anxiety* and *stress*), the structure and connectivity of the semantic network determines whether this activation dissipates or initiates a recursive cycle. This

occurs because highly central and connected nodes (often representing concepts with a high valence for the individual) activate neighboring nodes in a cluster with similar affective valence, reinforcing the same emotional patterns. Moreover, these secondary nodes activated, may have fewer connections and could potentially return activation to the starting node as well, reinforcing a long-lasting emotional loop characteristic of rumination.

Unlike neuroticism, conscientiousness, extraversion, agreeableness, and openness may serve as protective factors against mental health disorders [14,20]. Conscientious individuals have a goal-focused mindset, mitigating rumination. Extraversion correlates with positive affect and sociability, agreeableness promotes harmonious social interactions, while openness encourages creativity and flexible thinking. Collectively, these behavioral traits and thinking tendencies can foster dense positive clusters of nodes and weaken direct connections to negative concepts, potentially protecting from stress factors. Personality traits are thus studied both as a structural and as a behavioral moderator of emotional activation, potentially influencing which concepts are accessible, how strongly they interlink, and how stimuli are elaborated.

Psychometric scales as proxies for assessing mental health. To assess psychological well-being and its relationship with cognitive networks dynamics, three validated psychometric scales were employed: the Depression, Anxiety and Stress Scales (DASS-21), the Life Satisfaction Scale, and the Positive and Negative Affect Schedule (PANAS).

Depression Anxiety Stress Scales (DASS-21). The DASS-21 [23] is a widely used self-report scale, designed to assess the severity of three distinct psychological constructs: depression, anxiety, and stress. Each subscale consists of seven items that capture core symptomatology: depression (e.g. anhedonia, hopelessness), anxiety (e.g. hyperarousal, excessive worry), and stress (e.g., tension, irritability). By conceptualizing these constructs dimensionally, this tool is especially suited to examine how the different manifestations of distress operate in associative networks. In this context, we hypothesize that individuals with higher scores in the DASS-21 subscale scores may also exhibit cognitive networks with denser and more interconnected negative associations among nodes representing depression, stress and anxiety, much like found in other studies about math anxiety [10].

Life Satisfaction Scale. The Life Satisfaction Scale [24] measures global well-being by assessing individual's overall perception of life quality. The scale consists of five items that evaluate the extent to which individuals perceive their life as fulfilling and meaningful; unlike the DASS-21, this scale provides a cognitive evaluation of well-being, focusing on positive perception of well-being or its absence. In cognitive networks, individuals reporting greater life satisfaction may exhibit a network structure where positive concepts are more central or framed along other positive concepts, as found in past studies with the Emotional Recall Task [25].

Positive and Negative Affect Schedule (PANAS). The PANAS [26] is a psychometric scale that differentiates between positive affect (PA) and negative affect (NA). The PA subscale measures the frequency to which an individual experiences positive states (such as attention, determination, enthusiasm), while the NA subscale captures distress-related emotions such as fear, hostility, and guilt. From a network science perspective, high NA scores could be linked to stronger and persistent activation of negative semantic nodes, which in turn could reinforce distress, as found in past studies with the Emotional Recall Task [25].

The Emotional Recall Task (ERT). The ERT is a free-association paradigm designed to capture how individuals spontaneously retrieve and verbalize emotional experiences [27]. In its standard form, participants are asked to generate a fixed number of words—typically around ten—that describe how they have felt over a recent period (e.g., the past week or month). These self-reported words are then analyzed in terms of their affective properties (such as valence, arousal, and dominance) and their position within normative lexical networks. By mapping recalled words onto semantic networks, the ERT provides a window into the accessibility of emotion-laden concepts and how patterns of recall may reflect underlying psychological states, such as stress, anxiety, or depression. This approach has

been shown to link individual differences in recall with validated psychometric measures, making it a useful tool for exploring the relationship between emotional memory, well-being, and personality [28].

3. Manuscript Aims and Study Outline

Building on the methodological framework of cognitive network science [11], this work aims to extend the analysis of emotional recall patterns to two complementary studies. In Study 1, we investigate the relationship between the Emotional Recall Task (ERT) responses and psychometric measures of well-being in humans and LLMs. In Study 2, on the other hand, we focus on the link between ERT responses and personality traits. By doing so, three fundamental themes and research questions are addressed:

- In Study 1, we examine the extent to which human-generated associations correlate with clinical measures, such as the DASS-21 scale, the PANAS and the Life satisfaction scales. **RQ1:** Can spreading activation signals mirror psychological well-being indicators?
- In Studies 1 and 2, we compare and understand the ability of GPT-4, Haiku and Opus to simulate the associative patterns observed in humans. **RQ2:** Can these models can mirror human emotional dynamics under the lens of cognitive network science?
- In Study 2, we analyze the correlation between the results from recall tasks and the Big Five personality traits. **RQ3:** In either humans or LLMs, is there a relationship between the structure of associative memory and personality traits?

4. Materials and Methods

For Study 1, five distinct datasets were employed to examine the associations between words generated in the Emotional Recall Task and participants' psychometric outcomes as measured by the PANAS, DASS-21, and Life Satisfaction scales. For Study 2, five distinct datasets were employed to examine the associations between words generated in the Emotional Recall Task and participants' psychometric outcomes as measured by the PANAS, DASS-21, and Life Satisfaction scales. These datasets comprised two human participant datasets, which were sourced from previous studies [25,27,29], as will be explained later, and three datasets of LLM-simulated artificial participants, specifically GPT-4, Claude Haiku, and Anthropic Opus. This diversity in sources enables comparison of association patterns across different populations and the assessment of how closely artificial models reflect human-like representations of emotion and well-being.

For Study 2, we re-used data from a large-scale online survey conducted in the United States between May and August 2024, collected and re-shared on an Open Science Repository by De Duro and colleagues [29]. The survey was administered to a sample of 1,000 adult participants, who first completed a brief demographic questionnaire, followed by a series of psychometric assessments. Personality traits were measured using the IPIP-NEO Inventory (short form) [30], which evaluates neuroticism, extraversion, openness, agreeableness, and conscientiousness through five items per trait. After completing the scale, participants took part in the Emotional Recall Task (ERT), requiring them to freely generate words describing how they had felt in recent weeks. The resulting data used to construct individual activation trajectories, which were then analyzed to examine relationships between specific personality profiles and the activation strength of nodes associated with mental distress.

Human participants datasets Two datasets were used to analyze the relationship between emotional word associations and psychological states in human subjects. The first dataset, derived from a study by Ying et al. [27], contains data from 200 native English speakers recruited via Amazon Mechanical Turk. Participants provided ten emotional words describing their feelings over the past month, each accompanied by a self-rated frequency of experience, a valence rating (1–9 scale from unpleasant to pleasant), and an arousal rating (1–9 scale from calm to excited). These responses allowed for the computation of a valence-arousal emotional profile for each participant. After completing the emotional recall task, participants were administered several psychometric instruments: the Positive

and Negative Affect Schedule (PANAS) [26], the Depression Anxiety and Stress Scales (DASS-21) [23], the Satisfaction With Life Scale (SWLS) [24].

The second human dataset consists of Emotional Recall Task responses from a larger sample of 1000 individuals, collected by De Duro [29] to validate the association between personality traits and trust in LLMs using free-recall results. This dataset contains personality traits (the IPIP-NEO questionnaire for personality traits [31]) and is used to collect broader indicators of emotional and psychological functioning in order to generalize the activation results to a larger human sample. This dataset is used in both Studies 1 and 2.

Artificial participant datasets To simulate artificial participants, three distinct LLMs were employed: OpenAI's GPT-4, Claude Haiku 3.5, and Anthropic Opus 3.5. For each model, artificial profiles were generated using a fixed, standardized prompt designed to elicit emotionally relevant content and simulate answers to a psychometric questionnaire:

Impersonate a [x] years old [male/female/person].

Please use 10 English words to describe feelings you have experienced during the past month. Reply only with 10 words separated by a comma.

Please read each numbered statement and indicate how much the statement applied to you over the past week. The rating scale is as follows: 0 indicates it did not apply to you at all, 1 indicates it applied to you to some degree, or some of the time, 2 indicates it applied to you to a considerable degree or a good part of time, 3 indicates it applied to you very much or most of the time. Reply only with the vector number corresponding to your answers.

[Statements from the psychometric questionnaire y are listed.]

Repeat the two tasks independently [z] times.

Here, x represented an age value and ranged to match the age ranges in TILLMI by De Duro and colleagues [29], y was a questionnaire among DASS-21, Life Satisfaction and PANAS, z allowed to repeat 10 times the task, facilitating the simulations. Each LLM was tasked with independently generating hundreds of artificial participants by repeating the prompt across different fictional profiles. For each simulated participant, ten emotional words and full psychometric vectors were collected. These data were treated identically to those of human participants in subsequent analyses, allowing for a direct comparison of emotional word associations and activation dynamics across human and LLM-generated datasets. We excluded the IPIP-NEO Inventory from LLMs' simulations because the current literature indicates that LLMs without explicit prompting instructions might not possess clear, reliable and well-specified personality traits [32,33].

This design enabled us to assess the representational fidelity of LLMs in capturing emotional constructs and their relationship with mental health indicators. Furthermore, the use of three different models allowed for the identification of model-specific biases and performance differences in reproducing human-like semantic associations.

4.1. Preprocessing and Construction of the Network

Only participants whose entire set of recalled words from the Emotional Recall Task existed in the Small World of Words (SWOW) English dataset [17] were included in the simulations. This filtering step ensured that each word had a corresponding node within the associative network, preventing bias introduced by missing data. When misspellings occurred and a valid corrected version existed in the SWOW, the misspelled words were manually corrected; otherwise, the participant's full set was excluded to preserve consistency in network structure and input quality.

The cognitive network was constructed from the SWOW dataset [17], with nodes representing individual words and edges reflecting associative strengths based on lexical recall frequencies. This network was implemented using the NetworkX library (available at <https://networkx.org/>, Last Accessed: 06/10/2025).

For the dataset including PANAS and DASS-21 scales, participants' psychometric scores were linked to their *total network activation outputs*, i.e. the sum of all activation levels of specific nodes over time, to analyze possible relationships between psychological distress and emotional concepts spreading in memory.

4.2. Spreading Activation Dynamics

We adopted spreading activation as implemented in SpreadPy [10]. SpreadPy <https://github.com/dsalvaz/SpreadPy> (Last Accessed:26/09/2025) is a Python library for simulating spreading activation on single- or multi-layer cognitive networks. Each node i at time t carries an activation energy $e_{i,t}$, which updates in discrete steps according to a retention parameter r and a transfer function $\varphi(i \rightarrow j)$ that distributes residual energy to neighbors. The general update rule is:

$$e_{i,t+1} = r e_{i,t} + \sum_{j \in N(i)} \varphi(j \rightarrow i), \quad (1)$$

where $N(i)$ is the neighborhood of node i . In case of unweighted connections, SpreadPy assigns equal probability to all neighbors:

$$\varphi(i \rightarrow j) = \frac{(1-r) e_{i,t}}{\deg(i)}, \quad (2)$$

where $\deg(i)$ is the degree of node i . In the weighted case, transitions depend on normalized edge weights α_{ij} :

$$\varphi(i \rightarrow j) = (1-r) e_{i,t} \alpha_{ij}, \quad \text{with } \sum_j \alpha_{ij} = 1. \quad (3)$$

By tuning r and α_{ij} , SpreadPy provides a flexible framework for modeling how activation flows through semantic networks, making it a suitable tool for investigating memory biases, rumination, and emotional recall. Figure 1 contains an overview of the spreading activation dynamics on a cognitive network.

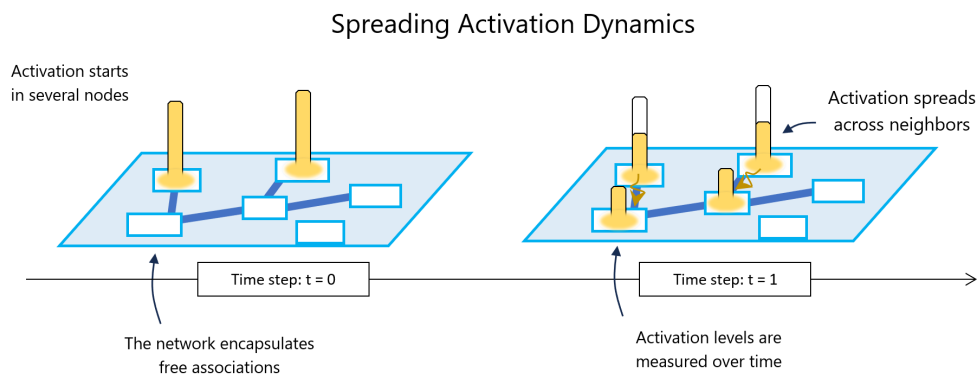


Figure 1. Visualization of the spreading activation dynamics over a given network topology and across subsequent time steps.

4.3. Spreading Activation Model Implementation

The simulation framework with SpreadPy [10] was designed to trace the semantic activation of the emotional concepts *stress*, *anxiety*, and *depression* within a cognitive network, following the recall of emotional states. Initially, all ten emotional words recalled by each participant, and expressed from the ERT data, were activated simultaneously. Then, activation spreading was run to simulate how other associated concepts mediated the flow of activation across the cognitive network of free associations, i.e. memory recall patterns.

This approach allowed the observation of the total activation level achieved by a given individual for a target concept related to mental health, i.e. one among "anxiety", "stress" and "depression". In other words, our aim was to identify how individual differences in reported emotions (and, in the case of the first dataset, psychological profiles) influence the dynamics of semantic activation toward negative emotional concepts.

In Study 2, the spreading activation framework described in Study 1 was applied to examine the relationship between personality traits and semantic activation dynamics. The goal was to assess whether individual differences in traits such as neuroticism, conscientiousness, or extraversion influenced semantic activation towards emotionally salient concepts like *stress*, *anxiety*, and *depression* following word recall. As in Study 1, participants' recalled ERT words served as simultaneous activation inputs in the network, and activation levels for each of the three target nodes were tracked over 200 computational time steps.

4.4. Batch Simulations and Correlation Analysis

Simulations employed a batch processing approach adopting the implementation of spreading activation of SpreadPy [10]. SpreadPy implements Collins and Loftus' spreading activation in single-layer and multiplex networks. We here used the single-layer version. For each participant, all ten recall words were activated simultaneously in the semantic network, with activation spreading tracked over 200 time steps. Activation levels of the target nodes "stress," "anxiety," and "depression" were recorded at each timestep.

The resulting time series of activation levels for each participant were visualized using log-log plots, and the final activation levels were extracted. For participants in the MTurk dataset, these final activation scores were statistically correlated with their psychometric results from the DASS-21 (Stress, Anxiety, Depression subscales) and PANAS (Positive and Negative Affect) scales, as well as their scores from the Life Satisfaction scale.

Kendall's Tau coefficient was selected for correlation analysis for its suitability for detecting monotonic yet potentially nonlinear relationships between the shape of activation curves and psychological measures. The visual and statistical comparison aimed to investigate whether higher levels of emotional distress corresponded to faster or more intense activation of negative emotional concepts. All simulations and visualizations were implemented using Spreadpy, Seaborn, and Matplotlib, ensuring methodological continuity with Study 2.

5. Results

5.1. Study 1

The first set of results stems from the correlation analyses between the lexical activation levels of specific negative concepts (anxiety, stress, depression) and scores from the various psychometric scales. The reference measure is the DASS-21, divided into three distinct subscales (DASS Anxiety, DASS Stress, DASS Depression), with PANAS Positive and Negative subscales and Life Satisfaction scores considered as well to explore broader well-being associations. All plots, presented in Figures 3, 4 and 5, derived from the study were converted into log-log scale, both on the X and Y axes, for both human data and LLMs. Correlations were then computed on these log-transformed values using scatter plots with confidence intervals.

5.1.1. Correlations Between Node Activation and Mental Health Scales

Table 1 presents the correlations between maximum activation levels of keywords (e.g. anxiety) and their psychometric dimensions (e.g. DASS-21 anxiety score), when activation originated from the ERT data.

Table 1. Kendall’s τ correlations between activation of negative concepts and psychological scales.

Scale	Humans	GPT-4	Haiku	Opus
DASS-Depression vs "depression"	0.375***	−0.029	0.136**	0.036
DASS-Anxiety vs "anxiety"	0.225***	0.149*	0.015	−0.123
DASS-Stress vs "stress"	0.298***	0.152*	0.007	0.005
Life Satisfaction vs "depression"	−0.380***	−0.071	−0.155**	0.051
Life Satisfaction vs "anxiety"	−0.254***	−0.132*	0.007	−0.059
Life Satisfaction vs "stress"	−0.244***	−0.127*	0.007	−0.048
PANAS positive vs "depression"	−0.351***	−0.089	−0.137*	0.025
PANAS positive vs "anxiety"	−0.127***	−0.138*	−0.087	−0.071
PANAS positive vs "stress"	−0.149***	−0.131*	0.009	−0.064

Notes: Significance levels — * $p < .05$, ** $p < .01$, *** $p < .001$. “n.s.” = non-significant. Dashes (−) indicate measures not applicable to LLMs.

For human participants, activation of negative concepts in the spreading activation model significantly and positively correlated with their corresponding DASS-21 subscales. Depression activation showed the strongest correlation with DASS-Depression (Kendall’s $\tau = 0.375$, $p < .001$). Similarly, anxiety activation positively correlated with DASS-Anxiety (Kendall’s $\tau = 0.225$, $p < .001$), and stress activation correlated with DASS-Stress (Kendall’s $\tau = 0.298$, $p < .001$).

These findings confirm that individuals reporting higher levels of psychological distress also exhibit stronger, faster and more persistent activation of distress-related concepts in their cognitive networks, supporting the memory bias hypothesis in emotional processing. In contrast, LLMs exhibited inconsistent correlation patterns compared to human participants. Haiku showed a weak positive correlation between depression activation and DASS-Depression (Kendall’s $\tau = 0.136$, $p = .009$), but nonsignificant correlation for anxiety (Kendall’s $\tau = 0.015$, $p = .772$) and stress (Kendall’s $\tau = 0.007$, $p = .899$). Opus demonstrated an inverse correlation between anxiety activation and DASS-Anxiety (Kendall’s $\tau = -0.123$, $p = .057$), and non-significant correlation for depression and stress, indicating that its network structure does not reflect human-like emotional connectivity. GPT, despite showing a significant correlation with DASS-Stress (Kendall’s $\tau = 0.152$, $p = .026$), yielded nonsignificant results between depression activation and DASS-Depression, as well as for anxiety activation and DASS-Anxiety. The patterns observed with DASS-Stress results, however, make GPT’s networks the most similar structure to human participants’ ones. Beyond DASS-21, the relationship between negative activation and well-being measures (PANAS Positive and Life Satisfaction) was examined.

In humans, analyses of well-being indicators revealed consistent inverse associations between negative concept activation and both positive affect and life satisfaction. This is also reported in Figure 2. Specifically, activation strength was negatively correlated with PANAS Positive across all three domains, with the strongest effect observed for depression ($\tau = -0.3510$, $p < .001$), followed by stress ($\tau = -0.1490$, $p < .05$) and anxiety ($\tau = -0.1265$, $p < .05$). Similarly, higher activation of distress-related nodes was associated with lower scores on the Life Satisfaction scale. This effect was most pronounced for depression ($\tau = -0.3801$, $p < .001$), but was also robust for anxiety ($\tau = -0.2538$, $p < .001$) and stress ($\tau = -0.2444$, $p < .001$). Together, these findings indicate that individuals whose recall networks more strongly activate concepts of depression, anxiety, and stress tend to report reduced positive affect and diminished satisfaction with life, highlighting the sensitivity of the Emotional Recall Task to broader aspects of subjective well-being. Compared to humans, LLMs displayed way weaker correlations between total activation scores and PANAS/Life Satisfaction scores (cf. Table 1).

The direction of the above findings suggests that the absence of autobiographical experience in LLMs creates a disconnection between lexical activation and subjective emotional states; while demonstrating learned associations between depression and negative affects, their activation patterns are weaker and non-systematic.

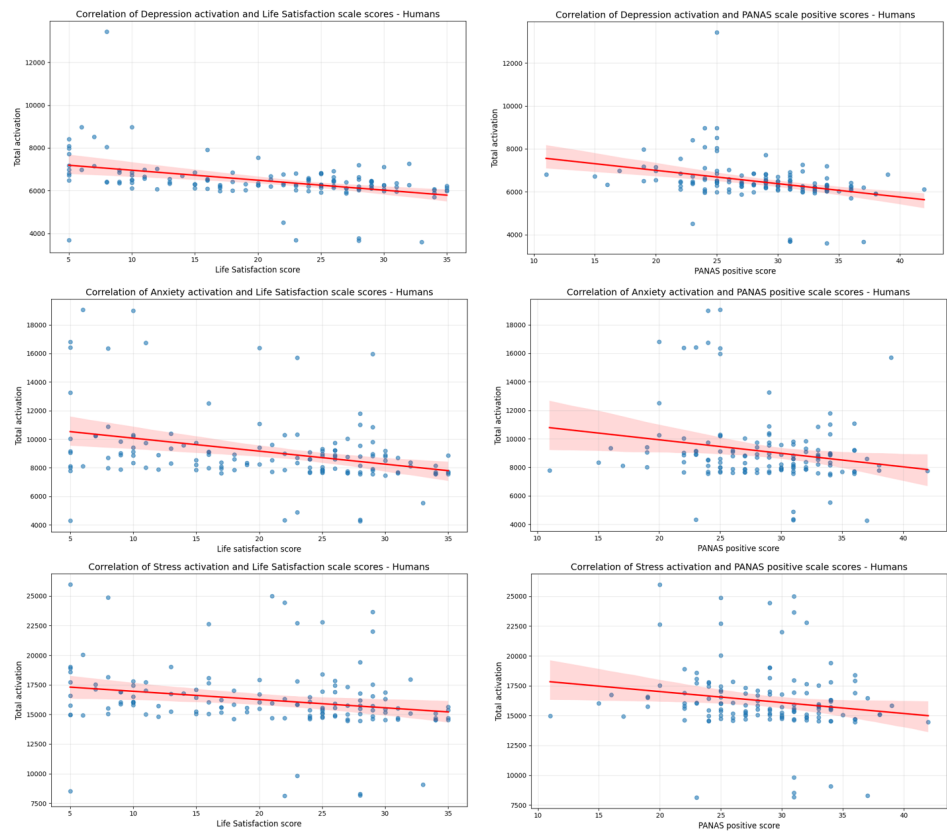


Figure 2. Scatter plots of total activation levels for concepts "anxiety", "depression" and "stress" for each individual human ERT data (representing a dot) versus the correspondent PANAS positive scores (right) and Life Satisfaction scores (left).

5.1.2. Differences Between Human Participants and LLMs

The log-log scale analysis revealed significant differences in activation intensity distributions between human participants and LLMs. First, human participants displayed greater variability compared to LLMs, with activation levels covering multiple orders of magnitude. For example, in cases where anxiety was triggered, some participants exhibited extreme spikes in activation, while others remained at lower, stable levels. This variability reflects individual differences in emotional processing, where psycho-social factors (e.g. personal memories, biases, and personality traits) modulate how strongly distress-related concepts are activated. In contrast, LLMs exhibited substantially less activation variability. Despite identical spreading activation mechanism for all models, extreme peaks of activation were rare, and the overall variability was compressed. Among the models, GPT-4 produced the most human-like activation curves, possibly due to its extensive training corpus and richer semantic representation. However, the absence of episodic memory prevented a full replication of the full range of activation intensities observed in human participants. Another key finding in the log-log curves is a pattern of convergence effect: after reaching peak activation, both human participants and LLM trajectories gradually decline and stabilize, reaching a metastable level where activation neither decreases and increases. For humans, this demonstrates functional universality; while individual differences shape the curves before reaching the maximum activation point, all human participants eventually reach the same levels of activation within their group, sharing approximately the same structures. LLMs exhibited a similar pattern, although with considerably less individual variability. This might be due to LLMs potentially showing emotional biases similar to humans, as already noticed in the relevant literature [34].

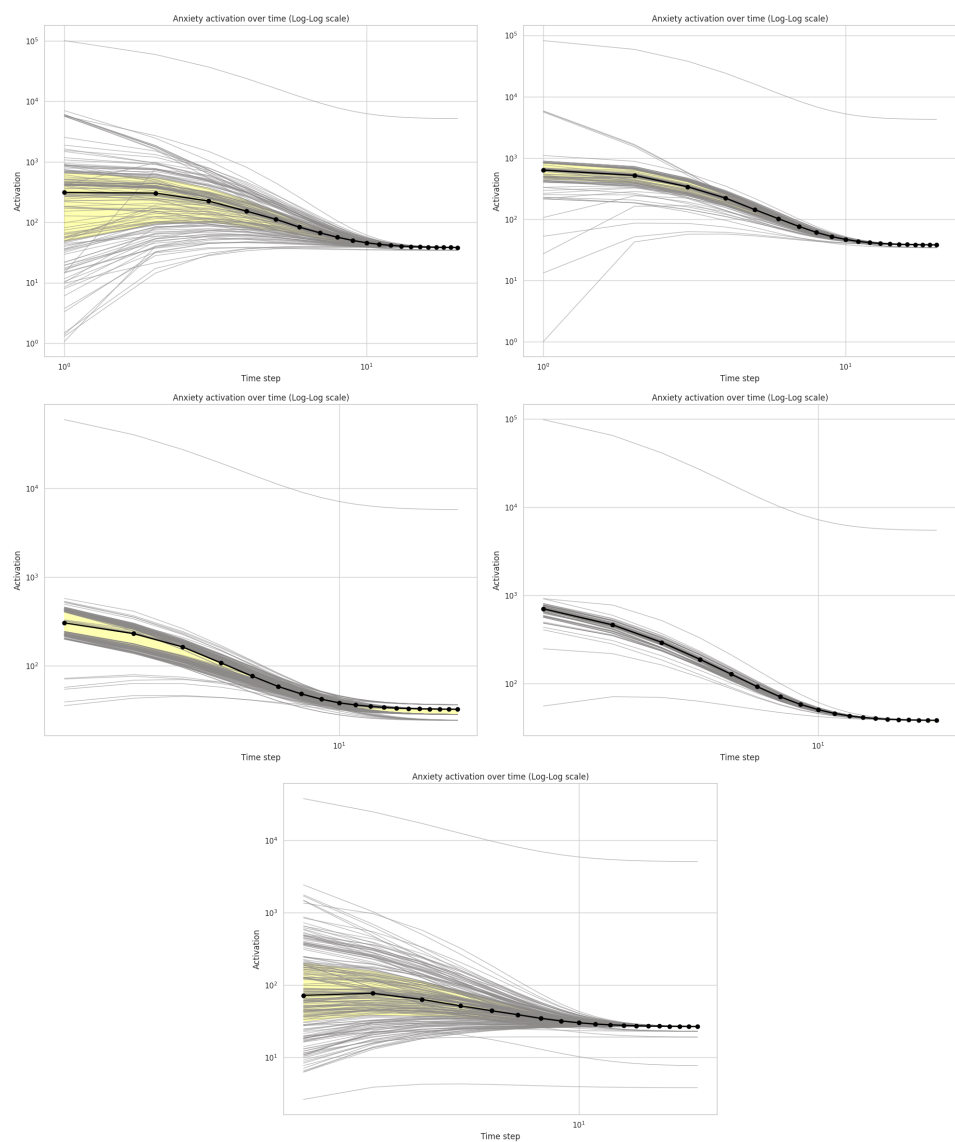


Figure 3. Time evolution across subsequent steps of activation levels for the word "anxiety" across free association networks when using: human ERT data from Study 1 (top left). GPT-4 ERT data (top right), Haiku ERT data (middle left), Opus ERT data (middle right) and human ERT data from Study 2 (bottom).

5.2. Study 2

Study 2 focuses on investigating correlations between personality traits and total activation levels of concepts like "anxiety", "depression" and "stress". Results are reported in Table 2 and refer only to humans. We did not consider simulations with LLMs in terms of personality traits because without explicit prompting instructions it is unclear to consider whether LLMs possess clear, reliable and well-specified personality traits, according to the literature [32,33].

Table 2. Kendall’s τ correlations between activation of distress-related concepts and Big Five personality traits in humans (values rounded to four decimals).

Trait	Stress	Anxiety	Depression
Conscientiousness	−0.0545	−0.0659	−0.1600**
Agreeableness	−0.0466	−0.0580	−0.1275*
Openness	−0.0466	−0.0580	−0.1275*
Extraversion	0.0032	0.0064	−0.1864***
Neuroticism	−0.0244	0.0046	0.1608**

Notes: Significance levels — * $p < .05$, ** $p < .01$, *** $p < .001$.

Humans with high neuroticism traits show stronger activation of concepts such as *anxiety* and *stress*, alongside higher peaks in activation for *depression*. Their tendency to worry and ruminate on negative experiences may explain the denser connectivity of their semantic network, facilitating rapid activation even without direct exposure to stressors. The correlation between neuroticism and depression activation (Kendall’s $\tau = 0.1608, p = .0013$), confirms a statistically significant association between higher neuroticism scores and increased activation of depressive concepts.

On the other hand, high extraversion seems to mitigate negative activation, as extraverted individuals demonstrate weaker activation of depression. The negative correlation between extraversion and depression activation (Kendall’s $\tau = -0.1864, < .001$) suggests that extraversion may act as a protective factor, reducing the spread of negative semantic activation, suggesting behaviors that could prevent ruminative loops of negative thoughts activation. Notably, this variability cannot be replicated in LLMs, as they do not possess personality traits; while they can statistically model semantic relationships between words, they lack the structural factors that shape personality. As a result, direct comparisons are not feasible.

These findings highlight both the potential of cognitive network science for understanding anxiety, stress, and depression.

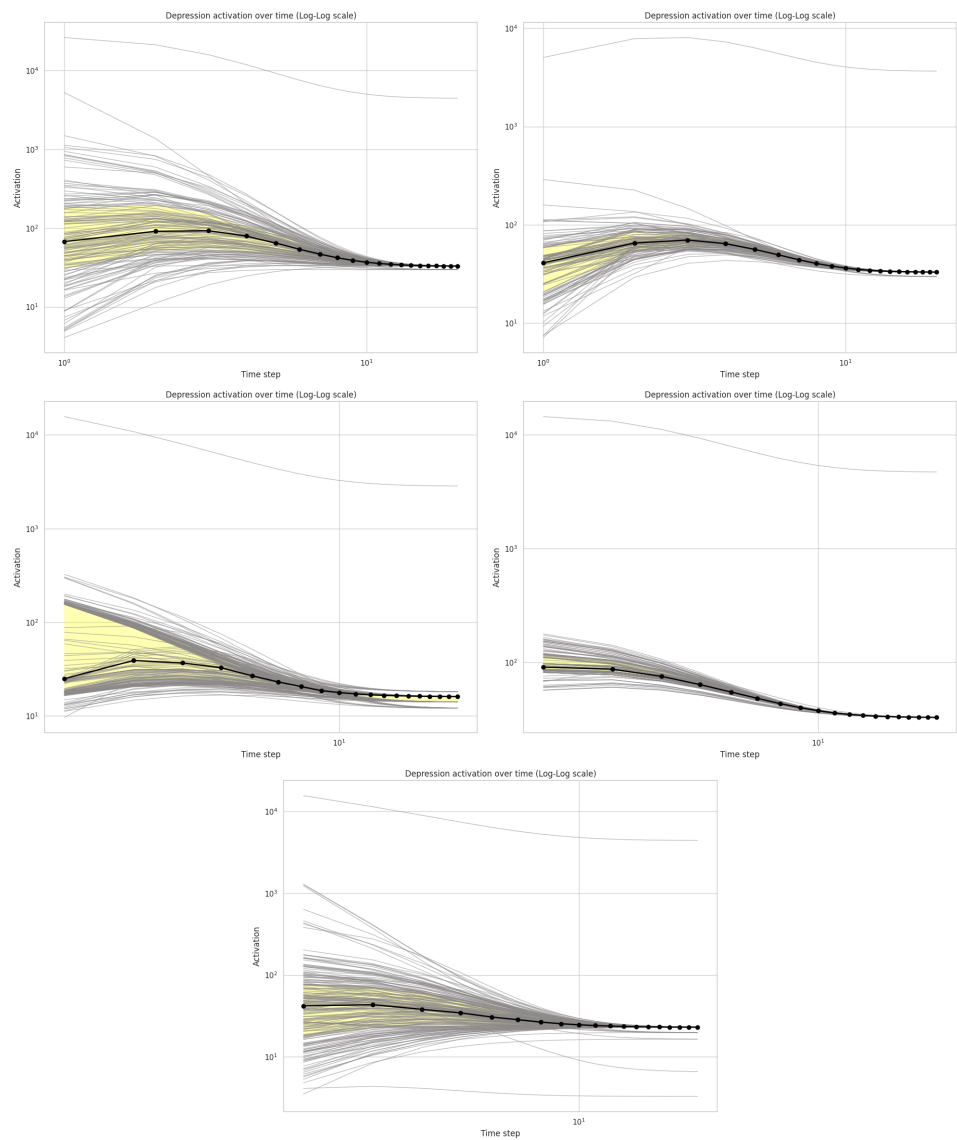


Figure 4. Time evolution across subsequent steps of activation levels for the word "depression" across free association networks when using: human ERT data from Study 1 (top left). GPT-4 ERT data (top right), Haiku ERT data (middle left), Opus ERT data (middle right) and human ERT data from Study 2 (bottom).

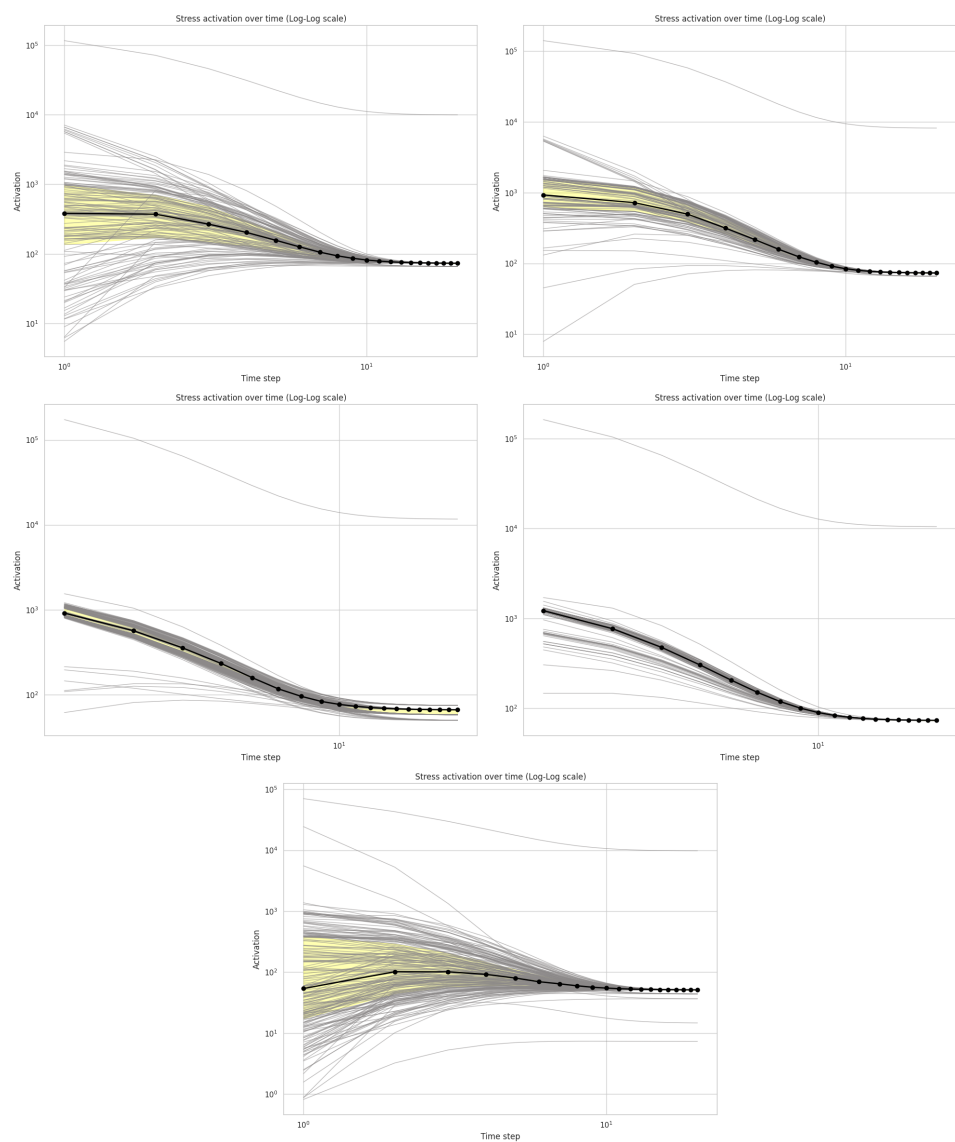


Figure 5. Time evolution across subsequent steps of activation levels for the word "stress" across free association networks when using: human ERT data from Study 1 (top left), GPT-4 ERT data (top right), Haiku ERT data (middle left), Opus ERT data (middle right) and human ERT data from Study 2 (bottom).

6. Discussion

This paper examined the relationship between the results from the Emotional Recall Task and mental health. By applying the spreading activation model to word pair lists generated through the ERT task, we hypothesized that individuals scoring higher on mental health scales (e.g., DASS-21) and those with higher neuroticism traits would exhibit stronger activation peaks of "cognitive energy" following initial activation of nodes concerning mental distress (e.g., stress, depression, anxiety). The findings supported these hypotheses, revealing stronger peaks of activation reaching these groups and with results comparable across humans and LLMs regarding the relationship between mental health scales' scores and word pairs generated from the ERT task. These results offer new insights into the structure and the functioning of emotional memory in its relationship with mental health.

Study 1 focused on the link between the word pairs generated from the Emotional Recall Task (ERT) and scores on psychometric scales such as the PANAS, DASS-21 and Life Satisfaction scales. We examined the possible correlation in both human participants and LLMs (Haiku, Opus and GPT-4) by generating simulated participants. This approach allowed us to investigate whether a relationship exists between psychometric scales and the ERT, and to what extent these associations correlate with clinical measures through simulations applied to ERT results and psychometric scales scores.

Study 1 also revealed recurrent patterns in activation trajectories among negative concepts, particularly in participants with higher distress scores. These loops likely reflect the ruminative circuits present in the individuals' cognitive structure, moving from one negative concept to the following ones. This quantitative pattern indicates that there are semantic closures in the cognitive networks that could lead to negative activation zones where activation is trapped, limiting exits to other clusters. This structure is not only consistent with models of rumination sustaining negative emotional states and some affective disorders, but also aligns with the "attractor states" theory proposed to study dynamic systems and applied to psychology. According to this theory, certain psychological states, thoughts and behavior become stable over time due to repeated reinforcement [35]. In the same way, negative clusters are automatically activated and reinforced, and this lexical activation is able to generalize to multiple situations over time, maintaining negative emotional states and thoughts and even supporting resistance to change and treatment in mental health disorders.

Additionally, Study 1 aimed at understanding whether LLMs are able to mirror the same emotional and mnemonic dynamics found in humans. The study's results demonstrated that the associations emerging from the ERT in humans are positively correlate with the results obtained from psychometric scales. Participants showing higher and faster activation peaks following words recall also scored higher on scales measuring negative affect (such as the DASS-21 and PANAS Negative) and lower on scales measuring positive affect (such as the Life Satisfaction Scale and the PANAS Positive). As shown also in past studies [1], LLMs showed widely varied performances: GPT-4 produced the most similar results to humans, though with lower activation peaks and less variability among simulated participants. Haiku and Opus, on the other hand, showed weaker and statistically nonsignificant correlations between ERT-generated activation levels and the scores of simulated participants across the different scales.

The findings from Study 1 indicate that lexical recall patterns, analyzed through the lens of associative networks and their cognitive activation dynamics can be successfully compared to scales designed to assess mental distress (such as the DASS-21, PANAS and Life Satisfaction scale), with cognitive activation towards words indicating emotional distress (such as *depression*, *anxiety* and *stress*) varying linearly with the scores obtained in these scales. This association confirms that emotional states are tightly linked to lexical memory access, aligning with previous research showing that negative emotions bias memory recall and information processing ([4,6]). Individuals with higher DASS-21 anxiety, depression, or stress scores showed significantly stronger, faster and more persistent activation of negative concepts. These distress-concept activation patterns suggest that emotional recall, assessed through free associations, can identify mental distress and is sensitive to the cognitive changes in thoughts patterns found in mental disorders [36,37].

Study 2 explored the associations between the Big Five personality traits and the activation of negative emotional concepts, specifically *anxiety*, *stress*, and *depression*, during a spreading activation simulation in cognitive semantic networks. The results revealed that high neuroticism scores correlated significantly and positively with increased activation of negative concepts, particularly depression, while high extraversion scores correlated negatively with depression activation. These results are consistent with the psychological literature identifying neuroticism as a key vulnerability factor for affective disorders, ([38,39] and extraversion as a potential protective trait from negative affective states [40]. These findings have implications on psychological theories, as lexical recall [9,10,41] could potentially be integrated into assessments for at-risk individuals: high centrality scores for *anxiety* or *depression* or high activation around certain nodes could signal a neuroticism personality trait, which in turn is associated with increased emotional vulnerability.

6.1. Memory in Humans vs. LLMs: The Role of Episodic Memory

The LLMs tested in Study 1 partly failed to replicate human responses, as Large Language Models' outputs did not exhibit the same associative or emotional patterns observed in human participants. This discrepancy aligns with past ones recently detected in cognitive psychology [42] and psycholinguistics [41]. In our mental health case, this discrepancy likely stems from the lack of autobiographical memory

[1,15]: Emotional states related to personal memory (as in the emotional recall task) also require autobiographical content, which is inherently absent in LLMs.

Study 1 involved the comparison of activation spreading dynamics between human participants and Large Language Models (LLMs). The comparison between the two groups highlighted both similarities and differences. While humans showed wide variability in activation levels across individuals, LLMs showed a sort of “compressed” distribution. This difference could be accounted for by the absence of episodic memory in LLMs, and also reflects the lack of autobiographical depth and emotional self-reference that characterizes human cognition and the memories that human form. Janik [43] claims that human memory integrates contextual, sensory and emotional features, resulting in richly personal and highly variable memory recall. In contrast, LLMs can only reproduce the statistical co-occurrence mechanisms and semantic regularities found within their learning corpora or due to their training processes [15].

Nonetheless, despite their lack of personal experience and memories [1], LLMs surprisingly displayed structured activation trajectories and certain regularities resembling those seen in human participants. This was especially true for GPT-4, whose spreading activation patterns most closely approximated those of humans, especially around the *stress* node activation. These similarities suggest that, although LLMs obviously lack the biological [44] and emotional [12,23] foundation of human cognition, their internal representations may partially reflect aspects of human-like semantic organization.

We highlighted how negative concept activation in humans was not only stronger but also more varied, hinting at a role of individual differences, potentially including personality traits, trauma exposure, stress levels or even presence or absence of symptoms of mental distress in modulating the recall patterns in their cognitive network. LLMs, in contrast, showed minimal variability within each model, lacking the heterogeneity that characterizes human emotional cognition, influenced by lived experiences. This again underscores the structural limitations of artificial cognition compared to human mental life. Raz and colleagues [45] show that LLMs show some similarities to human participants in terms of problem-solving and question complexity performance, warning about distinctive differences in the underlying cognitive processes between humans and LLMs. Mahowald and colleagues [46] emphasize that while humans and LLMs produce similar outputs in recall or language generation tasks, the underlying cognitive pipelines at work are still fundamentally different.

6.2. Limitations and Future Directions

Several limitations from this study must be addressed. First, while statistically significant, the correlations found were modest and should be interpreted cautiously; their effect sizes suggest that personality is merely one of many factors influencing semantic activation, and individual variability remains high. Second, the Emotional Recall Task [27] itself may be more sensitive to emotional states than traits, introducing noise into personality associations. While neuroticism is a quite stable trait [47], the activation of words concerning symptoms or negative concepts in one session may reflect momentary distress rather than trait-like tendencies. Therefore, multiple results over time from the same individuals would be necessary to better identify which results can be linked to personality traits and which to mental state influences. Third, as stated above, the link between mental health and personality traits is still unclear. While in literature some patterns have already emerged linking different personality traits to attitudes [48], thought patterns and emotional styles, researchers speak of tendencies rather than of clear, straight associations. Finally, although the study links personality types to semantic activation patterns, it does not examine the mechanisms through which this link occurs. Future studies could integrate measures of rumination [49] and other cognitive styles, like mind-wandering [44], to better specify these mediators and measure their impact on recall associations.

The strength of the correlations between activation and distress as measured by scales further highlights the possibility of using network-based tools in psychological assessment.

A major consideration emerging from this work concerns whether LLMs can effectively simulate patients with psychiatric symptoms. Despite LLMs being able to reproduce psycholinguistic data

with some effectiveness [50], the question of psychiatric symptoms is way more complex. Here, LLMs were able to partially generate responses that activated the same negative concepts targeted in human participants. However, here we found that the depth and variability of these activations diverged significantly. As described earlier, GPT-4 showed moderate alignment with human-like activation paths, particularly for the *stress* node, suggesting some capacity of the model to mimic emotional associations. However, its inconsistencies in correlating with *depression* or *anxiety* across different models reveals a sort of ceiling effect in their ability to simulate distress dynamics. Nevertheless, there is growing evidence that indicates that LLMs, when prompted appropriately, can mirror certain styles of thoughts [2,3,50] and, importantly, specific emotional tones. Both CounseLLMe by De Duro and colleagues [29] and other recent work by Wang and colleagues [51] both demonstrated that GPT-generated patient narratives and interactions well represent the communicative and emotional characteristics of real patients when guided with clinically relevant prompts. These results suggest that while the internal processes of LLMs are not rooted in personal experience, their training on emotionally charged corpora and on patient dialogues allows them to reconstruct the form of emotional expression as found in humans. Therefore, although the underneath mechanisms are different, the outputs produced by LLMs in certain conditions resemble those of human patients. This opens the way to novel intriguing directions for future works, making LLMs an interesting addition to clinical research and practice [52,53]. By embedding LLMs into therapeutic simulations for mental health trainees or diagnostic interviews, practitioners could explore differential emotional responses, identify linguistic markers of distress, and test hypotheses about psychopathology dynamics, provided their limitations are clearly understood [54].

7. Conclusions

The findings from both studies revealed that peaks of activation, indicating cognitive distress, correlate with higher scores on mental health scales and with the neuroticism personality type. Furthermore, this personality type has in turn been correlated to a higher risk of developing mental disorders, stressing the link between mental health and cognitive representations in semantic memory. This pattern, although more modestly, also emerged in LLMs, with GPT-4 showing the strongest alignment to humans. While not claiming precise mental state tracking from free recall tasks, this work demonstrated how cognitive network science can highlight mental health patterns of humans, potentially even identifying at-risk individuals, while also highlighting differences with Large Language Models.

Author Contributions: Conceptualization, A.C. and M.S.; methodology, A.C. and M.S.; software, A.C. and E.T.; validation, A.C. and E.T.; formal analysis, A.C.; data curation, A.C., E.T. and M.S.; writing—original draft preparation, A.C., E.T. and M.S.; writing—review and editing, A.C., E.T. and M.S.; visualization, A.C. and M.S.; supervision, M.S.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the COGNOSCO project, funded by University of Trento (grant ID *PS_{cal}227*).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new human data was generated for this study. The code for performing the activation simulations implemented in this work is available at the following repository (in the format of Google Docs working with Python): <https://osf.io/c7shb/>.

Acknowledgments: The authors acknowledge Enrico Perinelli and Tiziano Gaddo for valuable feedback at the early stages of this project.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. De Duro, E.S.; Improtà, R.; Stella, M. Introducing CounseLLMe: A dataset of simulated mental health dialogues for comparing LLMs like Haiku, LLaMAntino and ChatGPT against humans. *Emerging Trends in Drugs, Addictions, and Health* **2025**, *5*, 100170.
2. Cau, E.; Pansanella, V.; Pedreschi, D.; Rossetti, G. Selective agreement, not sycophancy: investigating opinion dynamics in LLM interactions. *EPJ Data Science* **2025**, *14*, 59.
3. Ashery, A.F.; Aiello, L.M.; Baronchelli, A. Emergent social conventions and collective bias in LLM populations. *Science Advances* **2025**, *11*, eadu9368.
4. Brosch, T.; Scherer, K.; Grandjean, D.; Sander, D. The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly* **2013**. <https://doi.org/10.4414/smww.2013.13786>.
5. Shields, G.S.; Sazma, M.A.; Yonelinas, A.P. The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews* **2016**, *68*, 651–668. <https://doi.org/10.1016/j.neubiorev.2016.06.038>.
6. Joormann, J.; Quinn, M.E. COGNITIVE PROCESSES AND EMOTION REGULATION IN DEPRESSION: Review: Cognitive Processes in Depression. *Depression and Anxiety* **2014**, *31*, 308–315. <https://doi.org/10.1002/da.22264>.
7. Mendl, M. Performing under pressure: stress and cognitive function. *Applied Animal Behaviour Science* **1999**, *65*, 221–244. [https://doi.org/10.1016/S0168-1591\(99\)00088-X](https://doi.org/10.1016/S0168-1591(99)00088-X).
8. Anderson, J.R. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior* **1983**, *22*, 261–295. [https://doi.org/10.1016/S0022-5371\(83\)90201-3](https://doi.org/10.1016/S0022-5371(83)90201-3).
9. Siew, C.S.Q. spreadr: An R package to simulate spreading activation in a network. *Behavior Research Methods* **2019**, *51*, 910–929. <https://doi.org/10.3758/s13428-018-1186-5>.
10. Citraro, S.; Haim, E.; Carini, A.; Siew, C.S.; Rossetti, G.; Stella, M. SpreadPy: A Python tool for modelling spreading activation and superdiffusion in cognitive multiplex networks. *arXiv preprint arXiv:2507.09628* **2025**.
11. Stella, M.; Citraro, S.; Rossetti, G.; Marinazzo, D.; Kenett, Y.N.; Vitevitch, M.S. Cognitive modelling of concepts in the mental lexicon with multilayer networks: Insights, advancements, and future challenges. *Psychonomic Bulletin & Review* **2024**, *31*, 1981–2004.
12. Kenett, Y.N.; Anaki, D.; Faust, M. Investigating the structure of semantic networks in low and high creative persons. *Frontiers in human neuroscience* **2014**, *8*, 407.
13. Citraro, S.; Vitevitch, M.S.; Stella, M.; Rossetti, G. Feature-rich multiplex lexical networks reveal mental strategies of early language learning. *Scientific Reports* **2023**, *13*, 1474.
14. Naragon-Gainey, K.; Watson, D. What Lies Beyond Neuroticism? An Examination of the Unique Contributions of Social-Cognitive Vulnerabilities to Internalizing Disorders. *Assessment* **2018**, *25*, 143–158. <https://doi.org/10.1177/1073191116659741>.
15. Stella, M.; Hills, T.T.; Kenett, Y.N. Using cognitive psychology to understand GPT-like models needs to extend beyond human biases. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2312911120.
16. Collins, A.M.; Loftus, E.F. A spreading-activation theory of semantic processing. *Psychological Review* **1975**, *82*, 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>.
17. De Deyne, S.; Navarro, D.J.; Perfors, A.; Brysbaert, M.; Storms, G. The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods* **2019**, *51*, 987–1006. <https://doi.org/10.3758/s13428-018-1115-7>.
18. Mohammad, S.M.; Turney, P.D. Crowdsourcing a word–emotion association lexicon. *Computational intelligence* **2013**, *29*, 436–465.
19. McCrae, R.R.; Costa, P.T. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology* **1987**, *52*, 81–90. <https://doi.org/10.1037/0022-3514.52.1.81>.

20. Ka, L.; R, E.; K, W.; G, J.; Lje, B. Associations between Facets and Aspects of Big Five Personality and Affective Disorders: A Systematic Review and Best Evidence Synthesis. *Journal of Affective Disorders* **2021**, *288*, 175–188. <https://doi.org/10.1016/j.jad.2021.03.061>.
21. Perkins, A.M.; Arnone, D.; Smallwood, J.; Mobbs, D. Thinking too much: self-generated thought as the engine of neuroticism. *Trends in Cognitive Sciences* **2015**, *19*, 492–498. <https://doi.org/10.1016/j.tics.2015.07.003>.
22. Nolen-Hoeksema, S.; Wisco, B.E.; Lyubomirsky, S. Rethinking Rumination. *Perspectives on Psychological Science* **2008**, *3*, 400–424. <https://doi.org/10.1111/j.1745-6924.2008.00088.x>.
23. Lovibond, P.; Lovibond, S. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy* **1995**, *33*, 335–343. [https://doi.org/10.1016/0005-7967\(94\)00075-U](https://doi.org/10.1016/0005-7967(94)00075-U).
24. Diener, E.; Emmons, R.A.; Larsen, R.J.; Griffin, S. The Satisfaction With Life Scale. *Journal of Personality Assessment* **1985**, *49*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13.
25. Stella, M.; Swanson, T.J.; Teixeira, A.S.; Richson, B.N.; Li, Y.; Hills, T.T.; Forbush, K.T.; Watson, D. Cognitive Networks and Text Analysis Identify Anxiety as a Key Dimension of Distress in Genuine Suicide Notes. *Big Data and Cognitive Computing* **2025**, *9*, 171.
26. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* **1988**, *54*, 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>.
27. Li, Y.; Masitah, A.; Hills, T.T. The Emotional Recall Task: Juxtaposing recall and recognition-based affect scales. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2020**, *46*, 1782. Publisher: American Psychological Association.
28. Stella, M.; Swanson, T.J.; Teixeira, A.S.; Richson, B.N.; Li, Y.; Hills, T.T.; Forbush, K.T.; Watson, D. Cognitive Networks and Text Analysis Identify Anxiety as a Key Dimension of Distress in Genuine Suicide Notes. *Big Data and Cognitive Computing* **2025**, *9*, 171.
29. De Duro, E.S.; Veltri, G.A.; Golino, H.; Stella, M. Measuring and identifying factors of individuals' trust in Large Language Models, 2025. Version Number: 2. <https://doi.org/10.48550/ARXIV.2502.21028>.
30. Johnson, J.A. Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality* **2014**, *51*, 78–89. <https://doi.org/10.1016/j.jrp.2014.05.003>.
31. Maples-Keller, J.L.; Williamson, R.L.; Sleep, C.E.; Carter, N.T.; Campbell, W.K.; Miller, J.D. Using item response theory to develop a 60-item representation of the NEO PI-R using the International Personality Item Pool: Development of the IPIP-NEO-60. *Journal of personality assessment* **2019**, *101*, 4–15.
32. Jiang, H.; Zhang, X.; Cao, X.; Breazeal, C.; Roy, D.; Kabbara, J. PersonaLLM: Investigating the ability of large language models to express personality traits. *arXiv preprint arXiv:2305.02547* **2023**.
33. Song, X.; Gupta, A.; Mohebbizadeh, K.; Hu, S.; Singh, A. Have large language models developed a personality?: Applicability of self-assessment tests in measuring personality in llms. *arXiv preprint arXiv:2305.14693* **2023**.
34. Cau, E.; Failla, A.; Rossetti, G. Bots of a Feather: Mixing Biases in LLMs' Opinion Dynamics. In Proceedings of the International Conference on Complex Networks and Their Applications. Springer, 2024, pp. 166–176.
35. Spencer, J.P.; Austin, A.; Schutte, A.R. Contributions of dynamic systems theory to cognitive development. *Cognitive Development* **2012**, *27*, 401–418. <https://doi.org/10.1016/j.cogdev.2012.07.006>.
36. Kaplan, D.M.; Palitsky, R.; Carey, A.L.; Crane, T.E.; Havens, C.M.; Medrano, M.R.; Reznik, S.J.; Sbarra, D.A.; O'Connor, M. Maladaptive repetitive thought as a transdiagnostic phenomenon and treatment target: An integrative review. *Journal of Clinical Psychology* **2018**, *74*, 1126–1136. <https://doi.org/10.1002/jclp.22585>.
37. Bemme, D.; Kirmayer, L.J. Global Mental Health: Interdisciplinary challenges for a field in motion. *Transcultural Psychiatry* **2020**, *57*, 3–18. <https://doi.org/10.1177/1363461519898035>.
38. Kotov, R.; Gamez, W.; Schmidt, F.; Watson, D. Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin* **2010**, *136*, 768–821. <https://doi.org/10.1037/a0020327>.
39. Ormel, J.; Jeronimus, B.F.; Kotov, R.; Riese, H.; Bos, E.H.; Hankin, B.; Rosmalen, J.G.; Oldehinkel, A.J. Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clinical Psychology Review* **2013**, *33*, 686–697. <https://doi.org/10.1016/j.cpr.2013.04.003>.
40. Sarubin, N.; Wolf, M.; Giegling, I.; Hilbert, S.; Naumann, F.; Gutt, D.; Jobst, A.; Sabaß, L.; Falkai, P.; Rujescu, D.; et al. Neuroticism and extraversion as mediators between positive/negative life events and resilience. *Personality and Individual Differences* **2015**, *82*, 193–198. <https://doi.org/10.1016/j.paid.2015.03.028>.

41. Vitevitch, M.S. Examining Chat GPT with nonwords and machine psycholinguistic techniques. *PLoS One* **2025**, *20*, e0325612.
42. Binz, M.; Schulz, E. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2218523120.
43. Janik, R.A. Aspects of human memory and Large Language Models, 2023. Version Number: 3, <https://doi.org/10.48550/ARXIV.2311.03839>.
44. Chang, M.; Sorella, S.; Crescentini, C.; Grecucci, A. Gray and White Matter Networks Predict Mindfulness and Mind Wandering Traits: A Data Fusion Machine Learning Approach. *Brain Sciences* **2025**, *15*, 953.
45. Raz, T.; Reiter-Palmon, R.; Kenett, Y.N. Open and closed-ended problem solving in humans and AI: the influence of question asking complexity. *Thinking Skills and Creativity* **2024**, *53*, 101598.
46. Mahowald, K.; Ivanova, A.A.; Blank, I.A.; Kanwisher, N.; Tenenbaum, J.B.; Fedorenko, E. Dissociating language and thought in large language models. *Trends in Cognitive Sciences* **2024**, *28*, 517–540. Publisher: Elsevier, <https://doi.org/10.1016/j.tics.2024.01.011>.
47. McAdams, D.P.; Olson, B.D. Personality Development: Continuity and Change Over the Life Course. *Annual Review of Psychology* **2010**, *61*, 517–542. <https://doi.org/10.1146/annurev.psych.093008.100507>.
48. da Silva, B.B.C.; Paraboni, I. Personality Recognition from Facebook Text. In Proceedings of the Computational Processing of the Portuguese Language; Villavicencio, A.; Moreira, V.; Abad, A.; Caseli, H.; Gamallo, P.; Ramisch, C.; Gonalo Oliveira, H.; Paetzold, G.H., Eds., Cham, 2018; pp. 107–114.
49. Bernstein, E.E.; Heeren, A.; McNally, R.J. Reexamining trait rumination as a system of repetitive negative thoughts: A network analysis. *Journal of Behavior Therapy and Experimental Psychiatry* **2019**, *63*, 21–27. <https://doi.org/10.1016/j.jbtep.2018.12.005>.
50. Binz, M.; Akata, E.; Bethge, M.; Brändle, F.; Callaway, F.; Coda-Forno, J.; Dayan, P.; Demircan, C.; Eckstein, M.K.;  ltet , N.; et al. A foundation model to predict and capture human cognition. *Nature* **2025**, pp. 1–8.
51. Wang, R.; Milani, S.; Chiu, J.C.; Zhi, J.; Eack, S.M.; Labrum, T.; Murphy, S.M.; Jones, N.; Hardy, K.; Shen, H.; et al. PATIENT-Ψ: Using Large Language Models to Simulate Patients for Training Mental Health Professionals, 2024. Version Number: 3, <https://doi.org/10.48550/ARXIV.2405.19660>.
52. Guo, Z.; Lai, A.; Thygesen, J.H.; Farrington, J.; Keen, T.; Li, K. Large Language Models for Mental Health Applications: Systematic Review. *JMIR Mental Health* **2024**, *11*, e57400. <https://doi.org/10.2196/57400>.
53. Lawrence, H.R.; Schneider, R.A.; Rubin, S.B.; Matari , M.J.; McDuff, D.J.; Jones Bell, M. The Opportunities and Risks of Large Language Models in Mental Health. *JMIR Mental Health* **2024**, *11*, e59479–e59479. <https://doi.org/10.2196/59479>.
54. Hua, Y.; Na, H.; Li, Z.; Liu, F.; Fang, X.; Clifton, D.; Torous, J. A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine* **2025**, *8*, 230. <https://doi.org/10.1038/s41746-025-01611-4>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.