Ralf Simsel

DIRECTED NEUROPLASTICITY THROUGH ARTIFICIAL INTELLIGENCE:

EMERGEOS AS COGNITIVE INFRASTRUCTURE

Practical work

Tallinn 2026

CONTENTS

INTRODUCTION

A person's personality is largely the sum of their repeatedly entrenched cognitive cycles. Most of these cycles form by chance and without conscious choice: fears, avoidance patterns, emotional associations, the inner monologue, and the direction of attention strengthen with each passing day, without the person ever choosing them deliberately. In everyday language this is called the formation of habits. In neuroscientific literature it is called neuroplasticity, the ability of the nervous system to change its structure and function in response to experience (Pascual-Leone et al., 2005). The starting point of this work is the question of whether this process can be turned from accidental into deliberate. Whether a software system can make a person's cognitive cycles visible to them and support their conscious restructuring.

Existing self-development systems broadly fall into two groups. One offers information: articles, courses, books on psychology. The other offers symptom relief: meditation apps, journals, conversational agents for emotional support. What is missing is infrastructure, something that helps a person observe their own cognitive architecture over time, understand it, and consciously reshape it. This gap was the real motivation behind the development of EmergeOS.

This work describes EmergeOS, a working web-based cognitive platform available at emergeos.app and already in use with real users. The system is not a journaling application in the usual sense. Journal entries are the system's input, but its analytical core deals with long-term cognitive modelling: identifying recurring thought patterns, tracking emotional trajectories, evaluating value alignment, and creating structured behavioural experiments. Each component is designed to support a specific mechanism described in the neuroscience literature as memory reconsolidation (Schiller et al., 2010).

The work presents three research questions. The first concerns the theoretical foundation: which psychological and neuroscientific models offer a reliable framework for a platform of directed neuroplasticity? The second is a question of positioning: how does EmergeOS differ from existing artificial intelligence applications, both functionally and conceptually? The third is practical: which cognitive patterns can the finished system actually identify based on my own use, and where do its limits lie?

The theoretical framework rests on several layers. The neuroscientific basis comes from Hebb's (1949) classical postulate of synaptic plasticity and from Pascual-Leone

and colleagues' (2005) review of the mechanisms of adult brain plasticity. On this foundation rests the theory of memory reconsolidation, which Schiller and colleagues (2010) demonstrated in a now-classic study published in Nature: fear-related memories are not immutable, and can be reshaped lastingly through the combined effect of activation and contradictory experience. At the clinical level, cognitive-behavioural therapy (Beck, 1963; 1979) provides categories whose linguistic markers a language model can identify in text. Acceptance and commitment therapy (Hayes et al., 2006) introduces the value-based dimension. Self-determination theory (Ryan & Deci, 2000) offers a framework for assessing basic needs. The work on emotional granularity (Kashdan et al., 2015) explains why more precise emotional differentiation matters for mental health. The handbook edited by Bennett-Levy and colleagues (2004) provides the structure of behavioural experiments, which ties all of these theories together into a practical intervention.

Technically, the system rests on retrieval-augmented generation (RAG), an architecture in which a search through the user's earlier writing precedes the language model's output. Béchard and Marquez Ayala (2024) have shown that this architecture substantially reduces the risk of hallucination, an unavoidable requirement in a sensitive field.

The practical need follows from the limitations of existing applications. AI-based reflection applications fall into two groups: conversational agents oriented toward emotional support, such as Replika and Wysa, on one side, and applications focused on the analysis of single journal entries, such as Mindsera, on the other. Neither offers long-term tracking of cognitive trajectory, citation-based validation, or structured behavioural experiments for measuring belief change. EmergeOS attempts to fill that gap.

The structure of the work follows the logic of a practical project. The first chapter sets out the theoretical framework, centred on neuroplasticity and how it can be directed. The second chapter considers the role of artificial intelligence as the foundation of a cognitive platform. The third chapter compares EmergeOS with existing applications. The fourth chapter explains the methodology. The fifth chapter is the practical heart of the work, describing the completed system with screenshots. The sixth chapter presents results drawn from the author's own use and feedback from pilot testers. The work concludes with a summary, a list of references, and annotations in Estonian and English.

It is important to note what this work is not. It is not a study of clinical efficacy. EmergeOS does not diagnose, treat, or replace professional help. The system is an external cognitive mirror for the user, not a therapist. The aim of the work is to demonstrate that such a platform is feasible, that it rests on solid psychological and neuroscientific foundations, and that initial use gives reason to believe in the value of further development.

1. THEORETICAL FRAMEWORK

1.1. Neuroplasticity and memory reconsolidation

For a long time the scientific consensus held that the human brain develops up to a certain age and undergoes no significant changes after that. Childhood was thought to be the only period in which the brain is capable of structural acquisition. This view was typical in neuroscience until the end of the 20th century. Doidge writes in the introduction to his book about how deeply entrenched this assumption was, to the point that scientists defending the idea of plasticity met with strong resistance (Doidge, 2007; Pascual-Leone et al., 2005). The idea of brain plasticity began with the Canadian psychologist Donald O. Hebb (1949), who in his book *The Organization of Behavior: A Neuropsychological Theory* set out a hypothesis later called Hebb's postulate. Hebb proposed that if one neuron repeatedly activates another, the synaptic connection between them grows stronger.

Today Hebb's postulate is a cornerstone of neuroscience, the foundation on which the modern understanding of the brain is built. The human cortex is in continuous reorganisation throughout life. Its structure changes in response to learning, experience, and deliberate practice. This is the conclusion reached by Pascual-Leone, Amedi, Fregni, and Merabet in their synthesis published in the *Annual Review of Neuroscience*, where the authors presented two decades of evidence (Pascual-Leone et al., 2005). Doidge (2007), in *The Brain That Changes Itself*, gives a range of clinical examples showing that the brain is reshapable: post-stroke movement impairments lasting decades have been recovered through repeated targeted training, patients with acquired balance disorders have learned to maintain balance through a sensory substitution device, and patients with obsessive-compulsive disorder have managed to interrupt their repetitive thought cycles through deliberate redirection of attention. In every one of these cases the mechanism is the same: repeated activation reshapes the physical structure of the brain.

The discovery that neuroplasticity can be engaged with consciously and deliberately is the starting point of this work. Most of a person's neuroplastic activity happens by chance: a person reinforces their fears, avoidance patterns, and inner monologue with each passing day without making any conscious decision about it. Doidge (2007) and earlier researchers, however, have shown that the same mechanism

can also be used in the opposite direction, to consciously reshape harmful thought and behaviour patterns. To distinguish this possibility from accidental plasticity, this work uses the term *directed neuroplasticity*. Classical examples of directed neuroplasticity are cognitive-behavioural therapy, meditation, and rehabilitation after stroke.

The principle of neuroplasticity explains how patterns become entrenched, but not how entrenched patterns can be changed. Much as the adult brain was once thought to be fixed, long-term memory was once considered strong and stable. The mid-twentieth century saw the rise of memory consolidation theory, according to which a fresh memory passes from an unstable state into a stable long-term form. Once this process was complete, the memory was thought to be written in. There was no mechanism explaining how an old memory might become open again, until Schiller, Monfils, Raio, Johnson, LeDoux, and Phelps (2010) showed in a study published in *Nature* that the memory consolidation theory was mistaken.

The logic of Schiller's study is straightforward. Fear-related memories are updatable, but only within a specific time window and through a specific procedure. Subjects were first taught to associate a particular colour with a mild electric shock. The next day, the fear memory was activated by showing them the same colour, but this time without the shock. During the following six hours, the so-called reconsolidation window, they were re-trained with the same colour, again without the shock. The result was striking: the fear response in these subjects was extinguished completely and permanently, even when tested a year later. A control group, given the same retraining outside the six-hour window, retained the fear response. The mechanism works in three steps: the memory must first be activated, then exposed to contradictory information, and all of this must happen within the six-hour window.

This finding has a direct meaning for the present work. It explains why purely intellectual understanding does not change emotional reactions. When a person tells themselves “I know my fear of failure is exaggerated,” they do not actually activate the fear memory at the emotional level, and it therefore remains unchanged. Lasting change requires two consecutive steps: first the activation of the belief, for instance by formulating a prediction (“if I try, I will fail”), and then a real experience that contradicts that prediction. The handbook of cognitive therapy edited by Bennett-Levy, Butler, Fennell, Hackmann, Mueller, and Westbrook (2004) formalises this principle as the structure of a behavioural experiment: identify a belief, formulate a prediction, take a real step, compare the outcome to the prediction. Their clinical conclusion is clear:

behavioural experiments produce stronger and more lasting belief change than purely verbal re-evaluation. This is where the neuroscientific finding and clinical practice meet.

1.2. Metacognition and the limits of self-analysis

Directed neuroplasticity presupposes the ability to observe one's own cognitive processes. The concept used for this ability is metacognition, or thinking about thinking, which means the human ability to notice, evaluate, and direct one's own thinking, learning, and reactions. One of the simplest examples of metacognition is the thought “I'm going to fail.” Without metacognition a person may take this as fact and slide into anxiety. With metacognition the person can notice: “I had the thought that I might fail. That does not necessarily mean I actually will.” Put simply, metacognition gives a person a certain distance between their thoughts and themselves. The person no longer automatically believes everything they think or feel, but can ask themselves about its origins. This also creates room to calm down, see the situation more clearly, and act more wisely. The concept of metacognition was introduced in the 1970s by John H. Flavell (1979). Research has linked metacognition with better learning ability, better decision-making, and better mental wellbeing (Nelson & Narens, 1990).

Metacognition also has limits. It helps a person notice individual thoughts or emotions, particularly when they are written down in a journal or some other form of self-expression. Frattaroli's (2006) meta-analysis of 146 studies confirmed that expressive writing has a meaningful effect on psychological wellbeing, especially when it is spread out over a longer period. At the same time, both Frattaroli's and Pennebaker's studies show that writing individual thoughts and feelings down in a journal does not automatically mean a person can also read longer-term patterns out of them. In other words, writing produces valuable raw material, but seeing the patterns and changing one's thinking requires that the material also be analysed.

Because a person does not always see the recurring patterns in their own thinking, metacognition can be supported by an external mirror: a therapist, a structured journal, a mood-tracking tool, or some AI-based analytical system. Such solutions do not necessarily replace a person's self-understanding, but they help to make visible the connections that go unnoticed inside any single moment.

1.3. The journal as a tool for detecting thought patterns

For a long time the journal has been a place where people structure their thoughts and express their emotions. At first glance its value seems to lie in the venting of feeling, but Ullrich and Lutgendorf (2002) showed that the actual benefit comes from somewhere else. When a person puts their experience into language, they are forced to structure it. A sentence cannot be left half-finished. Events must follow one another. Feelings need to be named. It is this work, what the brain does during the act of writing, that is beneficial, not what gets written down. Pennebaker (1997) and Frattaroli's (2006) large-scale meta-analysis confirm that regular writing on emotionally charged topics improves psychological wellbeing in a measurable way, particularly when the writing is distributed over a longer period.

A journal, however, remains within one person's perspective. Someone who has been writing for six months that “I can't handle this” may not notice on their own that the same sentence recurs every two weeks after a difficult meeting. They see today's entry and perhaps last week's, but not the pattern that becomes visible only across a comparison of ten entries. To see that pattern, you would need something that remembers every entry and can compare them. A therapist does this, but meets the person once a week and sees only what the person tells them, not what they write in their journal.

This is exactly the gap an AI-based journal fits into. Its difference from an ordinary conversational agent is significant. An ordinary chatbot can respond intelligently in the moment, but every conversation starts from zero because it has no memory of earlier sessions. An AI-based journal stores every entry and analyses them cumulatively: if the same thought pattern repeats over three different weeks, the system can see it and reflect it back to the user. Kim, Bae and colleagues (2024) studied the impact of such a journaling application on 28 patients diagnosed with depression over four weeks and found that it supported regular writing and helped clinicians better understand patients' day-to-day thinking between visits. Neshaei and colleagues (2025) demonstrated in a controlled study that AI-supported writing produces deeper reflection than independent writing.

1.4. Evidence-based psychological models

An AI-based journal must know what to look for in a person's text and how to interpret the patterns it finds. For this it uses psychological models. The selection of evidence-based psychological models proceeded from two criteria. The first criterion was clinical evidence: each chosen theory must have gone through controlled studies and had its effectiveness confirmed in independent systematic reviews. The second criterion was suitability for machine recognition: the theory must provide formally defined constructs that a language model can identify from the user's free text. In other words, there must exist a set of words and expressions that indicate a particular psychological pattern. The second criterion is decisive for EmergeOS, because many clinically effective theories, for instance psychodynamic therapy or dialectical behaviour therapy, do not offer the same level of linguistically operationalised categories. Cognitive-behavioural therapy was chosen primarily because the categories of cognitive distortions are formally defined and their linguistic markers are directly usable in machine analysis, as the use of large language models for detecting CBT distortions in text analysis demonstrates (Jiang et al., 2024). Acceptance and commitment therapy was added alongside CBT because it addresses the dimension of values and avoidance, which CBT does not cover. The work on emotional granularity provides a text-based model for emotion analysis. Self-determination theory adds a dimension for assessing the user's basic needs, which the other three theories do not address. Other evidence-based approaches, such as dialectical behaviour therapy, mindfulness-based cognitive therapy, and positive psychology, are clinically effective but do not meet the second criterion as well, or overlap functionally with the dimensions covered by the chosen four.

1.4.1. Cognitive-behavioural therapy

The origins of cognitive-behavioural therapy (CBT) reach back to Aaron T. Beck's 1960s research on the causes of depression. Beck (1963) noticed that the thinking of depressed patients repeatedly contained systematic errors, which he called cognitive distortions. These are automatic patterns of thought that seem rational at first but rest on faulty logic. As a simple example, if a person sends a message and a friend does not reply within an hour, the thought “they're angry with me” may seem like a logical conclusion. In fact this is the distortion known as mind-reading, in which a person infers another's internal state without evidence, when in reality the friend is simply working or did not notice the message. Beck originally distinguished six main distortions; later CBT handbooks have extended the list to fifteen or twenty categories. David, Cristea,

and Hofmann (2018) summarised the view that CBT is presently the evidence-based gold standard in the treatment of several psychiatric disorders.

In the context of this work, CBT is not used as a therapeutic method but as a linguistic framework for analysing text. The strength of cognitive distortions from the standpoint of text analysis is that they have typical linguistic markers. Black-and-white thinking manifests itself in absolute words such as “always,” “never,” and “nothing.” Catastrophising can be recognised in the construction “this means that,” followed by an extremely negative prediction (“I didn't answer the question, this means I'll fail the exam”). “Should” statements appear innocent at first (“I should be doing more”) but contain a rigid internal rule that turns ordinary fluctuation into guilt. EmergeOS detects these patterns systematically, and for every detection a specific quotation from the user's text is surfaced. It is this citation-based validation that distinguishes the system from its competitors.

1.4.2. Acceptance and commitment therapy

Acceptance and commitment therapy (ACT) belongs to the newer, so-called third wave of cognitive-behavioural therapy. Hayes, Luoma, Bond, Masuda, and Lillis (2006) define the central concept of ACT, psychological flexibility, as the ability to remain in contact with the present moment and to act in the direction of one's values even when uncomfortable thoughts and feelings are present. The difference from classical CBT is fundamental. The goal of CBT is to reshape negative thoughts. If a person thinks “I'm going to fail,” a therapist helps them reframe that thought more realistically. ACT starts from a different premise: fighting one's thoughts strengthens their hold. Instead, the person learns to notice their thoughts but not to let them dictate their action. The thought “I'm going to fail” may remain, but the person still takes the exam, because doing so is consistent with their values.

In the context of EmergeOS, the most directly applicable component of ACT is the values dimension. The system analyses which areas of life the user actively engages with in their writing and how, and surfaces the areas where the engagement is negative, avoidant, or sparse. If, for instance, the user writes often about work but almost never about relationships, that may be a signal of avoidance in an area that is nevertheless important to them. This is not a diagnosis or a claim that the person should be doing something differently. It is a reflective observation that gives the user the chance to judge for themselves whether the pattern in their writing matches their actual values.

1.4.3. Emotional granularity

Emotional granularity is a concept from the research group of Lisa Feldman Barrett that describes a person's ability to distinguish their feelings more precisely (Barrett et al., 2001). A person with low granularity uses general terms to describe their feelings, like “bad,” “upset,” or “stressed.” A person with higher granularity can distinguish more specifically: whether the feeling is frustration, shame, sadness, anger, or anxiety. This is not merely a linguistic difference. It shapes how the person regulates their emotions. When a person feels “bad,” it is hard to know what to do about it. When the person realises that what they feel is shame about a specific situation, a clearer path opens for engaging with that feeling. Kashdan, Feldman Barrett, and McKnight (2015) synthesised studies showing that people with higher emotional granularity use more appropriate strategies for regulating their emotions and experience lower levels of anxiety and depression symptoms.

EmergeOS analyses the user's text for emotions on three levels. The first level asks which emotion is present (for instance, “shame” or “frustration”). The second level evaluates the emotion's valence, whether it is positive or negative, and its arousal level, whether the state is calm or intense. The third level asks what role this emotion plays in the user's behaviour in the given context. The purpose of this structure is to help the user gradually formulate their inner world more precisely and, in doing so, train their granularity.

1.4.4. Self-determination theory

Self-determination theory (SDT) is a theory of motivation developed by Edward L. Deci and Richard M. Ryan. Its central idea is that a person's psychological wellbeing depends on the satisfaction of three basic needs: autonomy, competence, and relatedness (Ryan & Deci, 2000). Autonomy is the experience that one's activity is one's own free choice, not external pressure. A student who studies for an exam because the subject genuinely interests them experiences autonomy. A student who studies only so that their parents will not be displeased does not experience the same thing. Competence is the experience of being effective in one's actions, the sense that the person is capable of achieving what they set out to do. Relatedness is the sense that the person is meaningfully connected to others and that their relationships are supportive. A chronically unmet need damages motivation and wellbeing.

EmergeOS evaluates the three basic needs from the user's writing rather than through self-rating questionnaires. The system looks for substantive cues in the text. The dimension of autonomy is addressed by sentences that speak of control by others or perceived compulsion (“I have to,” “I don't get to choose”). Competence is addressed by descriptions of competence or incompetence (“I don't know how to do this well enough,” “I managed it”). Relatedness is addressed by mentions of relationship quality, descriptions of loneliness, or, on the other hand, observations of support from close people. The goal of the evaluation is not to diagnose but to make visible how much and how the user engages each area in their writing. This may reveal, for instance, that the person speaks often about achievement and work but almost never about relationships, which may point to unmet relatedness needs.

2. ARTIFICIAL INTELLIGENCE AS COGNITIVE INFRASTRUCTURE

2.1. The capabilities of large language models in text analysis

Until the early 2020s, automated psychological analysis of text was limited to what could be specified in formal rule sets. A system might look for the word “always” or “never” and infer the presence of black-and-white thinking on that basis. This approach broke down whenever a person expressed the same idea differently, for instance “every time I try, I fail.” The arrival of large language models fundamentally changed the situation. The model does not look for specific words in the text. It is able to grasp the meaning and structure that lies beneath different phrasings.

This does not mean a language model replaces a psychologist. The model is suited to pattern detection, not to therapeutic work. Omar and colleagues' (2024) systematic review of language-model applications in psychiatry shows that they are capable of supporting screening and linguistic analysis but remain limited when faced with complex clinical cases. Karkosz and colleagues (2024) conducted a randomised controlled trial with the Polish-language CBT-based conversational agent Fido and found a statistically significant reduction of anxiety and depression symptoms in the treatment group. EmergeOS uses the model for a single specific task: identifying linguistic markers in the user's text that correspond to categories defined in psychological theory, and then validating each detection against the original text. This is an engineering task, not a therapeutic one.

2.2. Retrieval-augmented generation and long-term memory

A large language model has one fundamental limitation: it has no memory from earlier sessions. If a user tells the model on Wednesday about a thought, the model on Thursday remembers nothing of it. This is not a defect but an architectural feature. For a journal-based system, however, it is a problem. Directed neuroplasticity assumes that the system sees a pattern over the course of months, not a single entry.

The solution to this problem is called retrieval-augmented generation (RAG). The principle is simple: before the model says anything in response to a user's entry, the system searches the user's earlier writings for similar passages and feeds them to the model as input. The model does not need to remember anything itself, because the

system brings it the necessary context every time. Béchard and Marquez Ayala (2024) showed that this architecture substantially reduces the model's tendency to hallucinate, that is, to produce statements that seem plausible at first glance but are factually wrong.

In EmergeOS this works as follows. Each journal entry is converted into a machine-readable vector (a set of 1536 numbers) that carries the semantic content of the entry. These vectors are stored in the database. When the user writes a new entry, the system searches the vector space for the ten most semantically similar earlier entries and feeds them to the model as input for analysis. In addition, the system also searches by temporal proximity: entries from the past two weeks. These two searches are weighted in a 55:45 ratio, preserving both contextual depth and immediacy.

Figure 1. EmergeOS's hybrid retrieval: the system combines temporally close entries and semantically similar entries weighted 55:45.

This is the conceptual difference between EmergeOS and an ordinary conversational agent. An ordinary conversational agent analyses a single message. EmergeOS analyses the user's entire body of writing over time.

2.3. Traceable reflection and ethical limits

In mental health, hallucination is especially risky. If a model tells a user that their writing shows signs of depression where none in fact exist, the result can damage the person's self-perception. Casu and colleagues' (2024) synthesis emphasises that AI applications in mental health must meet strict requirements for trustworthiness, privacy, and ethical limits. The system must not offer diagnoses, replace professional help, or give medical advice.

EmergeOS's answer to this is traceable reflection. The principle is that every output of the system must be traceable to a specific sentence in the user's text. If the system says "I see a mind-reading pattern in your writing," it must also show on the basis of which sentence the conclusion was drawn. The user can then judge for themselves whether the interpretation is correct.

This principle is realised in three mechanisms. First, extraction and interpretation are separated into two sequential stages, which reduces the model's load in any single pass. Second, a separate stage validates every output against the original text: if the system cites a sentence the user did not in fact write, the analysis is marked as faulty and

is not shown to the user. Third, strict output formats are used, which prevents the model from responding in free form.

On ethical limits, the system has clear exclusions. EmergeOS does not give diagnoses, does not use pseudoscientific frameworks (such as horoscopes or MBTI), does not assert causation without evidence, and does not imitate empathy. The aim is not to replace a therapist but to provide a transparent reflective observation.

3. COMPETITIVE ANALYSIS

3.1. Existing applications and their limitations

The competitive analysis in this work focuses on three applications that are closest to EmergeOS: Replika, Wysa, and Mindsera. These represent three different approaches to AI-based reflection, and the limitations of each at the same time point to EmergeOS's position.

Replika was founded in 2017 by Luka Inc. and has more than ten million users worldwide. The core idea of the application is to offer the user emotional support and companionship through a personalised conversational agent. Replika remembers user preferences and develops its conversational style over time, but its memory system is directed toward shaping the agent's persona, not toward analysing psychological patterns. Maples, Cerit, Vishwanath, and Pea (2024) conducted a study of 1006 university students on patterns of Replika use. Participants were lonelier than average but at the same time perceived high levels of social support. A notable share of users perceived the application as a friend, a therapist, and an intellectual mirror at once. Replika's limitation from the standpoint of a cognitive platform is clear: it is emotional companionship, not an analytical system.

Wysa is a CBT-based conversational agent developed by the company Touchkin eServices. Inkster, Sarda, and Subramanian's (2018) first real-world data evaluation showed a meaningful reduction in depression symptoms among highly engaged users over two weeks. A later randomised controlled trial confirmed a statistically significant decline in depression and anxiety indicators over four weeks (Leo et al., 2024). Wysa's main limitation lies not in its effectiveness but in its architecture: the application analyses each conversation separately, without long-term semantic memory. Every session essentially starts from zero. This is the memoryless-system problem addressed in section 2.2.

Mindsera is the closest functional competitor to EmergeOS. The application allows the user to keep a journal with AI analysis: for each entry the user receives feedback identifying emotions, cognitive distortions, and recurring themes. In 2025, Mindsera added a memory-profile feature that builds a longer-term view of the user's goals, relationships, and habits from their entire body of writing. This is the closest approach so far among AI journaling applications to long-term analysis. Mindsera's limitation,

however, is that its analysis does not tie its detections to specific quotations from the user's text and that it has no structured behavioural experiments for testing beliefs.

3.2. The market position of EmergeOS

EmergeOS differs from the three competitors mentioned in three ways, all of which follow from the system's position as a cognitive platform rather than a reflection application.

The first difference is the type of memory. Replika, Wysa, and Mindsera are architecturally synchronous: they primarily analyse a single conversation or entry. EmergeOS works with a long-term dataset and tracks the development of patterns over the course of weeks or months. This difference matters for neuroplasticity. Memory reconsolidation (Schiller et al., 2010) requires repeated activation and contradiction over a longer time interval. The analysis of a single entry cannot capture recurrence, which is the core of directed neuroplasticity.

The second difference is evidence-based grounding. The three competitors detect distortions to some extent, but none of them ties the detection to quotations from the user's own text. The validation stage of EmergeOS checks each detected distortion against the original text. If the cited sentence does not match the original, the analysis is marked as faulty. The user therefore sees exactly which sentence the analysis is based on. This is the practical implementation of the principle of traceable reflection.

The third difference is the presence of behavioural experiments. Replika and Wysa offer advice and exercises, but these are not tied to a specific identified belief, nor do they measure belief change after the experiment. EmergeOS's experiments contain a clear statement of the belief, a prediction, and a pre- and post-rating scale. This structure follows directly from the Bennett-Levy et al. (2004) handbook and rests on the principle of memory reconsolidation.

Taken together, these differences give EmergeOS a position that can be described as follows: it is not a competitor to Replika, Wysa, or Mindsera in the classical sense, because it does not belong in the same category. Replika is emotional companionship. Wysa is CBT-based symptom relief. Mindsera is an AI-analysed journal. EmergeOS is a cognitive platform whose purpose is to support long-term directed neuroplasticity. The system is publicly available at emergeos.app and is already in use with real users.

4. METHODOLOGY

This work combines several research and development methods, the interplay of which follows from the nature of the work itself. The subject is a software system that cannot be studied purely theoretically or purely empirically, and the development itself was a year-and-a-half-long learning process.

4.1. Research approach

This work is a theoretical-practical development project. The aim was not to conduct a classical empirical study with a large sample, measurements, and statistical analysis, but to study the theoretical foundations and on that basis design a prototype of an AI-based journal.

The work has two interconnected parts. The first part studied the theoretical background of neuroplasticity, metacognition, and selected psychological models. The second part used that knowledge to build the prototype of an AI-based journal. The methodology can therefore be described as a combination of literature analysis and prototype development.

The work does not directly measure neuroplastic change in the brain and does not claim to prove that an AI-based journal produces such change. Neuroplasticity is treated in this work as a theoretical framework that helps to understand why repeated self-reflection, noticing one's thoughts, and practising new ways of behaving may be connected to learning and change.

4.2. The research aim from a methodological perspective

The methodology follows from the general aim of the work: to explain how the principles of neuroplasticity, metacognition, and selected psychological models can be used in creating an AI-based journal prototype. Four main activities were carried out to reach this aim. First, the nature of neuroplasticity and metacognition was studied. Second, the possibilities of artificial intelligence for supporting metacognition and self-reflection were analysed. Third, the psychological models suitable for the logic of an AI-based journal were identified. Fourth, a prototype was developed that helps the user notice and track their own thoughts, emotions, and recurring patterns over time. The

focus of the work, then, is not on evaluating the impact of a finished solution but on building a theoretically grounded prototype.

4.3. Gathering and analysing theoretical material

In the first stage of the work, material was gathered and analysed in three main areas: neuroplasticity, metacognition, and AI-based self-reflection.

For neuroplasticity, the focus was on how the brain changes as a result of learning, repetition, experience, and the formation of habits. For metacognition, the focus was on how a person notices, monitors, and evaluates their own thinking, learning, and behaviour. For AI-based journaling, the focus was on how artificial intelligence could help the user structure their thoughts, ask clarifying questions, and identify recurring patterns.

Theoretical material was selected from peer-reviewed scientific articles in PubMed, ScienceDirect, Frontiers, and Nature. Primary scientific articles were preferred over secondary summaries. Alongside academic sources, more practically oriented material about AI-based journaling was also reviewed: documentation of existing applications (Replika, Wysa, Mindsera), user studies, and recent conference papers (for instance Kim et al. 2024 on the MindfulDiary study at CHI). These sources gave a picture of how analogous solutions have been built in practice and what their limitations are.

4.4. The selection and evaluation of psychological approaches

One important part of the work was identifying which psychological approaches could be used in building an AI-based journal. The aim was not to provide a therapeutic tool but to find approaches whose principles could be used to support self-reflection, the noticing of thought patterns, and the tracking of patterns over time.

The selection of evidence-based psychological models proceeded from two criteria. The first was clinical evidence: each chosen theory had to have gone through controlled studies and had its effectiveness confirmed in independent systematic reviews. The second was suitability for machine recognition: the theory had to provide sufficiently formally defined constructs that a language model could identify from the user's free text. In other words, there had to exist a set of words and expressions that indicate a particular psychological pattern.

This second criterion is decisive for EmergeOS, because many clinically effective theories, for example psychodynamic therapy or dialectical behaviour therapy, do not offer the same level of linguistically operationalised categories. Cognitive-behavioural therapy was selected primarily because the categories of cognitive distortions are formally defined and their linguistic markers are explicitly usable in machine analysis, as the use of large language models for the detection of CBT distortions in text analysis shows (Jiang et al., 2024). Acceptance and commitment therapy was added alongside CBT because it addresses the values and avoidance dimension that CBT does not cover. The emotional granularity approach provides a text-based model for analysing emotions. Self-determination theory adds the dimension of assessing the user's basic needs, which the other three theories do not address. Other evidence-based approaches, such as dialectical behaviour therapy, mindfulness-based cognitive therapy, and positive psychology, are clinically effective but do not meet the second criterion as well or overlap functionally with the dimensions covered by the chosen four.

These approaches were not used in the work as clinical interventions but as frameworks that help the user make better sense of their inner world and behaviour. For example, the cognitive-behavioural approach fits the journal context because it helps distinguish situation, thought, feeling, and behaviour. The work on emotional granularity helps the user name their emotions more precisely. Self-determination theory makes it possible to make sense of how the needs for autonomy, competence, and relatedness may affect motivation and self-perception.

4.5. The prototype development process

As the practical and most substantive part of the work, a prototype of an AI-based journal was developed, accessible at emergeos.app. The aim of the prototype was to show how artificial intelligence could support a user's metacognitive self-reflection and help them notice recurring thought, feeling, and behaviour patterns. The development process drew on the principles of neuroplasticity, metacognition, and the selected psychological approaches addressed in the theoretical part of the work.

Before the substantive development work could begin, a range of technical skills had to be acquired. These were both AI-related and more generally information-technological. Specifically, the development process required learning to work with Google's machine learning course materials (vector representations, neural networks,

transformer models), prompt engineering for steering large language models, the use of Claude as the main development agent inside the agent-based development environment Claude Code, the use of the NestJS and React frameworks, the use of Supabase and its pgvector extension, and the building of integrations with various third-party services. These skills were acquired during practical development as each stage demanded, rather than through a separate prior learning cycle.

As the first substantive step, the general aim and use case of the prototype were defined. The journal was designed as a tool the user could use to analyse everyday or recurring situations. The central use case was a situation in which the user wants to better understand their reactions, thoughts, emotions, or behaviour patterns.

Second, the theoretical principles were tied to the possible functions of the journal. From the principles of metacognition, the journal had to help the user notice their thinking and evaluate their reactions. From the cognitive-behavioural approach, questions were added that help distinguish situation, thought, feeling, and behaviour. From the emotional granularity approach, questions were added that guide the user to name their feelings more precisely. From self-determination theory, the need to make sense of experiences relating to autonomy, competence, and relatedness was taken into account.

Third, the core logic of the journal was put together. The prototype was designed so that the user begins with a free-text entry, and the AI raises clarifying reflective questions on that basis. The aim of the questions is not to give the user judgements but to help them better structure their experience. The logic of the journal moves generally from description to analysis: first the situation is described, then thoughts and feelings, then behaviour, and finally possible patterns and next steps.

Fourth, an initial set of questions and instructions was developed. The questions were worded to support self-reflection without being diagnostic or therapeutic. For example, questions such as “What were you thinking in this situation?”, “Which feeling did you notice most strongly?”, “Have you experienced a similar reaction before?”, and “What might you try differently next time?” were used. Such questions help the user observe their inner processes and connect them to specific situations.

Fifth, the prototype's ability to track recurring patterns over time was designed. For this, the elements of journal entries that could be compared were defined: recurring situations, recurring thoughts, frequent emotions, typical reactions, and the user's own stated next steps. This structure supports the aim of the work, because it joins together

the metacognitive activity of monitoring one's own thinking and the theoretical idea, from neuroplasticity, of the possibility of noticing and changing recurring patterns.

Sixth, the initial technical solution of the prototype was built. The prototype was built as a working web application using React and NestJS, with Supabase as the database and Anthropic's Claude as the language-model core of the analysis. The aim was not to create a finished commercial product but to show how an AI-based journal could substantively function and what kind of interaction it should have with the user.

Seventh, autoethnography was applied, a qualitative research method in which the researcher uses their own experience as a data source to understand a phenomenon in a broader context (Ellis, Adams, & Bochner, 2011). In this work I used the prototype as the author of the research and as the developer of the technical solution over several months and documented observations systematically in a separate notes file. After each analysis I evaluated the accuracy of the identified distortions, my reaction to the tone of the analysis, and the structure of the experiments. In addition, the system automatically logs cases in which the validation stage rejects an analysis, which provides more objective data than personal observation alone.

The development process was therefore iterative: principles were derived from theory, questions and use logic were created on the basis of those principles, and these were then assembled into a working application. The result of the work is an initial concept of an AI-based journal that could be further tested with users, refined, and technically developed in future.

4.6. The foundations of prototype construction

The construction of the prototype followed the principles of neuroplasticity, metacognition, and the selected psychological approaches set out in the theoretical part of the work. The aim of the prototype was to create a solution that helps the user better notice their own thoughts, emotions, and behaviour patterns and track their recurrence over time.

The construction of the prototype was guided by the following principles. First, the journal must direct the user toward self-reflection rather than supplying ready-made judgements. Second, the AI's questions must help the user distinguish situation, thought, feeling, and reaction. Third, the journal must support the noticing of recurring patterns over time. Fourth, the wording used must be understandable to an ordinary

user. Fifth, the AI must not give diagnoses or replace a psychologist or therapist. Sixth, the prototype must rest on the chosen psychological approaches but use them in the form of self-analysis, not therapy.

This approach allowed the theoretical part of the work to be tied to a practical output. The prototype was not separately evaluated in this work through a user study or expert review, and is therefore treated as an initial developmental solution that could be tested more broadly and refined in future.

5. DESCRIPTION OF THE PROTOTYPE

5.1. The aim of the prototype

EmergeOS is a prototype of an AI-based journal whose aim is to help the user notice their recurring thought, feeling, and behaviour patterns and to track their change over time. The system is not a therapist and does not give diagnoses. It is an external mirror for the user, capable of seeing in their own writing what they themselves miss, because they are too close to the text.

The prototype is publicly available at emergeos.app and is already in use with real users. As the practical part of the work, this chapter addresses six main elements of the system: the user journey within the application, the journal's core logic and question system, pattern analysis, behavioural experiments, the long-term tracking view, and the crisis-detection mechanism.

5.2. The user journey within the application

The user journey in EmergeOS is a cycle that begins with opening the long-term view and ends with launching a new experiment, before returning to the beginning. It is not a linear process but a closed loop in which each pass makes the user's cognitive model of themselves more precise.

A typical user session begins from the progress view, which displays a comparison between the last month and the one before: which thought patterns have grown more frequent, which have decreased, how many entries have been made, and how many experiments have been completed. This is the application's meta view, giving the user a sense of where they are. From there they move to the journal to write a new entry. The entry becomes input for the system, and soon an analysis is produced that surfaces the identified patterns together with quotations from the user's own text. If the analysis reveals a need for behavioural intervention, the user moves to the experiments view, where they choose their next experiment. During the experiment the user rates the strength of their belief before and after, and describes the actual outcome. If needed, the user can converse with the system to make better sense of their experience. Each such cycle feeds the next analysis with more precise data.

5.3. The journal’s core logic and questions

The journal is the user's primary input point. Each entry has a date, a mood selection (Positive, Neutral, Low, Tense, Calm), and a free-text field. The system does not put a long questionnaire in front of the user to fill out. Instead, it offers soft guiding quick-prompt chips that help them begin when they do not know where to start. These chips are “What am I observing about myself?”, “What pattern is emerging?”, and “What shifted since my last entry?”. All three questions are oriented toward observation rather than emotional expression and support the metacognitive dimension of reflection (Flavell, 1979).

The most important element of the journal view, however, is the personalised question displayed in a green box, generated by the system from the analysis of the user's earlier entries. Figure 2 shows one such question the system suggested to a specific user on a specific day: “Ralf, when was the last time someone did or said something kind to you (even small), what did your body notice in that moment, and what did you do with that kindness afterward?” This question is not chosen at random. It is directly tied to the system's earlier observation that this user's writing under-represents the need for relatedness and that the user tends to discount positive experiences. The question is constructed to guide the user toward noticing what they usually let slide past.

Figure 2. The new journal entry view. The header contains the date and mood selection. The green box holds the personalised question generated by the system from the analysis of earlier entries. Below are three soft starting points as quick-prompt chips.

After an entry is made, it joins the user's journal log, which is a chronological list of all their entries. Figure 3 shows this view. Each entry has a date, a title, and a few-line excerpt. The user can search and sort entries and reopen any entry to read or expand it. The word count next to the title (“469w”) shows the length of the entry, which is also a signal of how deeply the user has gone into exploring their experience.

Figure 3. The journal list view. On the left is the chronological list of all entries with search and sort tools. On the right is the full content of a selected entry.

5.4. Pattern analysis and reflection

Journal entries do not remain as individual records. The system analyses them cumulatively: each new entry is linked to the context of earlier entries, and the analysis

works on the full body of writing rather than only on the most recent entry. The result of the analysis appears in the insights view, which is the working network of the user's cognitive profile. This view is divided into several parts: threads, tensions, identified thinking traps, cognitive loops, emotions and behaviours, and value alignment.

Figure 4 shows the threads and tensions section. A thread is a recurring theme or dynamic the system has identified in the user's writing. For instance, the “Performance-based self-worth loop” describes a pattern in which the user tends to interpret single mistakes or rest as evidence of their lesser worth. Each thread has a short description, evidence (either quotations from the user's text or a note when the current entry contains no direct evidence), and an expandable “why this matters” block. Tensions are dynamics in which there is a contradiction between two of the user's values or needs, for instance the need to achieve versus the need to rest.

Figure 4. The threads and tensions section of the insights view. Each thread is an identified recurring dynamic together with evidence from the user’s text and an explanation of why it matters.

The detection of thinking traps is a direct application of the CBT-based framework. Figure 5 shows the identified loop “Performance-Based Self-Worth Loop,” visualised as a sequence of steps: slip or rest (trigger) → automatic identity inference (labeling) → emotional response (shame/urgency) → behavioural response (compensatory achievement or avoidance) → short-term outcome (temporary relief or renewed urgency) → REPEATS. Each loop is marked with the specific cognitive distortions operating within it; in this case Labeling, “Should” statements, and avoidance. The same view also shows the identified emotions together with their valence (positive/negative) and arousal level (medium/high arousal), following the model of emotional granularity from Barrett and colleagues (2001).

Figure 5. Visualisation of an identified cognitive loop together with its steps, the associated distortions, and the emotions and behaviour patterns that follow.

The value alignment section evaluates the user's three basic needs from self-determination theory: autonomy, competence, and relatedness (Ryan & Deci, 2000). Figure 6 shows this section. Each need has a score from zero to ten and a short explanation of which signals in the user's writing have conditionally shaped that score. The system is honestly transparent about what its rating is based on: if a specific entry contains no direct quotations, the system flags this separately (for instance “no direct

quotes in this snapshot; inference drawn from recent pattern history"). This is the practical realisation of the principle of traceable reflection.

Figure 6. The value alignment section based on self-determination theory. The three basic needs are scored on a scale from zero to ten together with an explanation and a note on how much data the rating is based on.

5.5. Behavioural experiments

Identifying patterns on its own does not change anything. A user may consciously recognise their destructive patterns for years and still fall back into them repeatedly. It is precisely this gap between awareness and change that is the reason behavioural experiments exist. An experiment is not a task or a habit imposed on the user; it is a deliberate behavioural step whose aim is to create a specific contradiction with a specific identified belief. The handbook of cognitive therapy edited by Bennett-Levy and colleagues (2004) formalises this structure, and Schiller and colleagues' (2010) theory of memory reconsolidation explains why it works.

Figure 7 shows the full structure of the experiments view. On the left is a list of active experiments, each with a short note on which pattern it interrupts. The right panel shows a specific experiment in detail. In the example the experiment is "Two 10-minute sensory switch breaks across the week," whose target belief is "Short breaks or sensory shifts won't change my mood or the self-criticism, only big fixes matter." The pre-experiment belief is set at 57 percent. During and after the experiment the user uses a slider to rate how strongly they still believe this belief afterwards. The difference between the pre- and post-rating is the operational measure of belief change.

Figure 7. The behavioural experiments view. On the left is a list of active experiments with categories. On the right is the full structure of the selected experiment: target belief, pre-experiment rating scale, rating slider, and two open-ended questions for describing the experience.

The two text fields for completing the experiment matter. The first asks "What actually happened?", which forces the user to separate the actual experience from the prediction they expected. The second asks "What did you learn?", which is a metacognitive step in which the user connects the specific experience to their broader understanding of themselves. When the experiment is complete, the whole sequence is saved to the database: target belief, prediction, actual outcome, what was learned, and the difference between the pre- and post-rating. These data become the input for the next analysis.

5.6. Long-term tracking and the progress view

The value of a single analysis is limited. Real value emerges when the system can show how the user's patterns change over time. For this the application has a progress view that can be treated as the system's meta-analytical layer. Figure 8 shows this view. The upper section (Longitudinal Assessment) gives a numerical overview: how many insight analyses the system has generated, how many experiments have been completed (Behavioral Integration), how many journal entries the system tracks (Emotional Baseline Stability), and the current consistency streak.

The second section of the view, Pattern Recurrence, compares the last 30 days to the preceding 30 days. Each pattern is visualised as two bars: PREV (the previous period) and NOW (the current period). The user in the figure can see, for example, that “Should” statements have grown by 13 units in their writing compared with the previous month, mind-reading by 11 units, and catastrophising by 8 units. Such numbers are not themselves a judgement or a diagnosis. They are objective signals that make visible to the user which patterns they have strengthened over the last month.

Figure 8. The progress view. The upper section shows the main indicators of long-term assessment. Below is a comparison of the last 30 days with the previous 30 days, where each pattern is shown as two bars: the previous and the current period.

It is this view that distinguishes EmergeOS from an ordinary journal. An ordinary journal stores entries but cannot show whether the user's cognitive patterns strengthen or weaken across months. The AI-based analytical layer makes this possible and gives the user a dataset on which they can make more deliberate decisions about their own process.

5.7. Crisis detection

Crisis detection deserves separate attention. Because EmergeOS works with sensitive material and a user may write thoughts referring to self-harm or suicide, a three-layer detection mechanism is built into the system. This is not an additional feature but a fundamental design requirement that follows directly from the ethical standard for mental-health applications (Casu et al., 2024).

The detection mechanism combines a keyword search using bilingual risk terms, a sentiment-classification step, and combined escalation logic, which allows the system to distinguish four severity levels: low, medium, high, and critical. In the critical case,

normal analysis is halted entirely, location-sensitive crisis resources are displayed (in Estonia: peaasi.ee, 116 123, and emergency line 112), and the system clearly emphasises that this is an AI, not a substitute for a professional. This mechanism realises the ethical principle described in the methodology chapter, that the system must not give diagnoses or replace professional help.

6. RESULTS AND ANALYSIS

This chapter presents observations from several months of personal use, combined with feedback from pilot testers and metrics from the development log. The results should be treated as a qualitative signal of how the prototype functions, not as evidence of clinical efficacy.

The clearest positive result concerns the detection of linguistically marked cognitive distortions. Black-and-white thinking, characterised by absolute words such as “always,” “never,” and “nothing,” was detected consistently and accurately. Catastrophising, which surfaces in sentence constructions predicting the worst-case scenario, was also detected reliably, especially where the phrase “this means that” was followed by an extremely negative prediction. Mind-reading, the conviction about other people's intentions without evidence, was detected with more difficulty. This is to be expected, because the detection of mind-reading cannot rest on linguistic markers alone and requires contextual understanding. In such cases the system's confidence level remained low, which is methodologically correct behaviour.

The validation mechanism worked as designed. The development log showed that several analyses were rejected because the quoted sentences did not exactly match the source text. The model had reworded them in ways that changed the meaning. This is precisely the behaviour the validation step was designed to prevent. The result is consistent with research on RAG systems (Béchard & Marquez Ayala, 2024).

Emotion analysis was more accurate when my text contained clear emotional descriptions. When I wrote about events rather than feelings, the system had to infer emotions indirectly, which led to greater uncertainty. The result is consistent with the literature on emotional granularity: more precise analysis requires more precise input (Kashdan et al., 2015). Value alignment scores turned out to be one of the more accurate parts of the analysis. Areas I wrote about little and negatively received low scores. This is methodologically correct behaviour: the score does not measure the user's actual satisfaction but how the user reflects that area in their writing.

The system's main advantage over its competitors is contextuality. The analysis does not tell the user abstractly that they tend to catastrophise. It shows a specific sentence from an entry eight weeks ago that displays the same pattern as something written yesterday. A moment of recognition like this carries a different psychological weight from a general observation. Compared with Wysa, the greatest weakness is

speed. Wysa gives feedback immediately. EmergeOS needs at least three entries and two to five minutes of analysis time. To first-time users this may feel uncomfortably slow, but the design choice is deliberate, because depth takes time.

One important personal observation concerns the role of experiments. In the first weeks it seemed that the analysis itself was the system's main value. Later it became clear that the actual change happens through the experiments. The experiment “do a ten-minute micro-product test” produced stronger belief change than ten analyses all confirming the same pattern. This corresponds exactly to the predictions of Bennett-Levy and colleagues (2004) and confirms the practical importance of the memory reconsolidation mechanism. Awareness without a contradicting experience is not enough. The greatest value of EmergeOS does not lie in its analysis but in the experimental cycle that follows from it.

Three recurring themes emerge from the observations of pilot testers. The first concerns the use of quotations. The fact that the analysis pointed to specific sentences from the user's own text generated trust in a way that abstract analysis does not. Several testers phrased it in roughly this way: this is from my own words, the system is not asserting anything I did not write myself. Second, the pre- and post-rating scale of the experiments initially felt too technical but became more intuitive after the first session. Third, several testers were surprised by how precisely the system identified areas in which they actually felt unfulfilled. This supports the platform's core hypothesis that a person's own text contains signals they themselves do not notice.

CONCLUSION

This practical work addressed the development and functioning of the AI-based cognitive platform EmergeOS. The system is publicly available at emergeos.app and is already in use with real users. The work posed three research questions, which can now be answered.

The first research question concerned the theoretical foundation. A directed neuroplasticity platform requires a multi-layered framework. The neuroscientific foundation comes from Hebb's (1949) postulate of synaptic plasticity and Pascual-Leone and colleagues' (2005) review of adult brain plasticity. The theory of memory reconsolidation (Schiller et al., 2010) provides the specific mechanism by which cognitive patterns become open to updating, but only when they have been activated and meet contradiction. Cognitive-behavioural therapy offers linguistically defined categories through which to identify patterns (Beck, 1963; 1979). Acceptance and commitment therapy adds the value-based dimension (Hayes et al., 2006). Self-determination theory provides a framework for assessing basic needs (Ryan & Deci, 2000). The emotional granularity approach supports more precise mapping of the inner world (Kashdan et al., 2015). The behavioural experiments framework (Bennett-Levy et al., 2004) gives a specific mechanism for reshaping identified beliefs through reconsolidation. None of these foundations is sufficient on its own. Their combination is what forms the conceptual basis of the cognitive platform.

The second research question concerned functional differentiation. EmergeOS is not a competitor to Replika, Wysa, or Mindsera in the classical sense, because it does not belong in the same category. Replika is emotional companionship. Wysa is CBT-based symptom relief. Mindsera is an AI-analysed journal. EmergeOS is a cognitive platform that works with cumulative memory, ties each detection to a specific citation, and structures behavioural experiments for measuring belief change. This position opens a new market segment rather than competing in existing ones.

The third research question concerned the system's actual capability. Linguistically marked distortions, such as black-and-white thinking and catastrophising, are detected consistently. Patterns that require contextual understanding, and emotions inferred from indirect text, remain less certain. The validation mechanism functions as designed and detects reworded quotations. The most important personal observation is that the actual value of the system shows itself not in the analysis itself but in the behavioural

experiments that follow, which create the contradiction needed to engage the neuroplasticity mechanism.

The main limitation of the work is the small sample and the personal nature of the use, which does not allow conclusions about clinical efficacy. Further development is already moving in two directions. First, a broader user study is being prepared to evaluate the platform's effectiveness in a more diverse sample. Second, EmergeOS is already a scaling product at emergeos.app, the long-term effect of which will need to be evaluated through long-term data analysis.

Looking back at the question with which the work began, whether a software system can turn neuroplasticity from accidental into deliberate, the answer is conditionally affirmative. The system is able to detect linguistically marked patterns reliably and to do so in a way that the detected patterns are verifiable through specific quotations. Through the layer of experiments, the system can create contradicting experiences that engage the reconsolidation mechanism. The system does not see everything, and there are gaps, but the basic mechanism required for it to work is present and in real use. In that sense EmergeOS shows that such a system can be built and used in practice. Whether it reliably produces lasting cognitive change is a separate question, and answering it will require a broader and longer-term study than this work could provide.

REFERENCES

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, 15(6), 713–724.

Beck, A. T. (1963). Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, 9(4), 324–333.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.

Béchard, P., & Marquez Ayala, O. (2024). Reducing hallucination in structured outputs via Retrieval-Augmented Generation. *arXiv preprint* arXiv:2404.08189.

Bennett-Levy, J., Butler, G., Fennell, M., Hackmann, A., Mueller, M., & Westbrook, D. (Eds.). (2004). *Oxford guide to behavioural experiments in cognitive therapy*. Oxford University Press.

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). Evaluating Generative AI in Mental Health: Systematic Review of Capabilities and Limitations. *JMIR Mental Health*, 11, e62963.

David, D., Cristea, I., & Hofmann, S. G. (2018). Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. *Frontiers in Psychiatry*, 9, 4.

Doidge, N. (2007). *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. Viking.

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Forum Qualitative Sozialforschung*, 12(1), Article 10.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.

Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*, 132(6), 823–865.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley.

Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, 6(11), e12106.

Jiang, M., Yu, Y. J., Zhao, Q., Li, J., Song, C., Qi, H., Zhai, W., Luo, D., Wang, X., Fu, G., & Yang, B. X. (2024). AI-enhanced cognitive behavioral therapy: Deep learning and large language models for extracting cognitive pathways from social media texts. *arXiv preprint* arXiv:2404.11449.

Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2024). Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial. *JMIR Formative Research*, 8, e47960.

Kashdan, T. B., Feldman Barrett, L., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, 24(1), 10–16.

Kim, T., Bae, S., Kim, H. A., Lee, S.-W., Hong, H., Yang, C., & Kim, Y.-H. (2024). MindfulDiary: Harnessing large language model to support psychiatric patients' journaling. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Article 701.

Leo, A. J., Schuelke, M. J., Hunt, D. M., Miller, J. P., Areán, P. A., & Cheng, A. L. (2024). Effectiveness of a Mental Health Chatbot for People With Chronic Diseases: Randomized Controlled Trial. *JMIR Formative Research*, 8, e50025.

Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research*, 3, 4.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, 26, 125–173.

Neshaei, S. P., Rietsche, R., Su, X., & Wambsganss, T. (2025). Metacognition meets AI: Empowering reflective writing with large language models. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13601

Omar, M., Soffer, S., Charney, A. W., Landi, I., Nadkarni, G. N., & Klang, E. (2024). Applications of large language models in psychiatry: A systematic review. *Frontiers in Psychiatry*, 15, 1422807.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, 28, 377–401.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, 8(3), 162–166.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463(7277), 49–53.

Ullrich, P. M., & Lutgendorf, S. K. (2002). Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine*, 24(3), 244–250.

ANNOTATSIOON

Simsel, R. (2025). *Suunatud neuroplastilisus tehisintellekti abil: EmergeOS kui kognitiivne infrastruktuur*. Praktiline töö. Tallinna Vanalinna Hariduskolleegium.

Praktilise töö raames arendati välja toimiv veebipõhine kognitiivne platvorm EmergeOS, mille eesmärk on toetada kasutaja suunatud neuroplastilisust. Süsteem on kättesaadav avalikult aadressil emergeos.app ning töötab juba praegu reaalsete kasutajatega. Töö positsioneerib EmergeOS-i uue tootekategooriana, mis erineb olemasolevatest tehisintellektipõhistest refleksioonirakendustest kontseptuaalselt, mitte ainult tehniliselt.

Teoreetiline raamistik tugineb mitmele kihile. Neuroteaduslik alus tuleb Hebbi (1949) sünaptilise plastilisuse postulaadist ja Pascual-Leone jt (2005) ülevaatest täiskasvanu aju plastilisusest. Mälu taaskonsolideerimise teooria (Schiller jt, 2010) annab konkreetse mehhanismi tuvastatud uskumuste ümberkujundamiseks. Kliinilist alust pakuvad kognitiiv-käitumisteraapia (Beck, 1963; 1979), omaksvõtmise ja pühendumise teraapia (Hayes jt, 2006), enesemääratlemise teooria (Ryan ja Deci, 2000), emotsionaalse granulaarsuse käsitlus (Kashdan jt, 2015) ning käitumuslike eksperimentide raamistik (Bennett-Levy jt, 2004). Tehnilist arhitektuuri toetab täiendusppõhise genereerimise põhimõte (Béchard ja Marquez Ayala, 2024). Metoodikana kasutati autoetnograafilist lähenemist mitme kuu pikkuse isikliku kasutamise jooksul.

Tulemused näitavad, et süsteem tuvastab keeleliselt selgeid kognitiivseid moonutusi usaldusväärselt, kuid kaudsete emotsioonide tuletamine jääb nõrgemaks. Töö peamine empiiriline järeldus on, et süsteemi tegelik väärtus avaldub mitte analüüsis endas, vaid sellele järgnevates käitumuslikes eksperimentides, mis loovad mälu taaskonsolideerimise käivitamiseks vajaliku vastuolu. EmergeOS demonstreerib, et suunatud neuroplastilisus tarkvara abil on teostatav reaalsus.

Märksõnad: suunatud neuroplastilisus, kognitiivne platvorm, mälu taaskonsolideerimine, tehisintellekt, kognitiiv-käitumisteraapia, käitumuslikud eksperimendid, suured keelemudelid.

ABSTRACT

Simsel, R. (2025). *Directed Neuroplasticity through Artificial Intelligence: EmergeOS as Cognitive Infrastructure*. Practical Work. Tallinn Old Town Educational College.

This practical work developed a functioning web-based cognitive platform, EmergeOS, designed to support users in directed neuroplasticity. The system is publicly available at emergeos.app and is already used by real users. The work positions EmergeOS as a new product category, conceptually distinct from existing artificial intelligence based reflection applications rather than merely a technical variant of them.

The theoretical framework draws on multiple layers. The neuroscientific foundation comes from Hebb's (1949) postulate of synaptic plasticity and Pascual-Leone et al.'s (2005) review of adult brain plasticity. Memory reconsolidation theory (Schiller et al., 2010) provides the specific mechanism for restructuring identified beliefs. Clinical grounding is provided by Cognitive Behavioural Therapy (Beck, 1963; 1979), Acceptance and Commitment Therapy (Hayes et al., 2006), Self-Determination Theory (Ryan & Deci, 2000), research on emotional granularity (Kashdan et al., 2015), and the behavioural experiments framework (Bennett-Levy et al., 2004). The technical architecture is grounded in retrieval-augmented generation (Béchard & Marquez Ayala, 2024). The methodology is autoethnographic, based on several months of personal use.

Results indicate that the system reliably identifies linguistically marked cognitive distortions, while inferring emotions from indirect descriptions remains weaker. The main empirical conclusion is that the system's actual value lies not in the analysis itself but in the behavioural experiments that follow, which create the contradiction required to trigger memory reconsolidation. EmergeOS demonstrates that directed neuroplasticity through software is a feasible reality.

Keywords: directed neuroplasticity, cognitive platform, memory reconsolidation, artificial intelligence, cognitive-behavioural therapy, behavioural experiments, large language models.