PsychExtract: Literature Review

Foundations of Insight and Emotion in Psychotherapy

Insight as a Therapeutic Mechanism

Insight is widely recognised as a core driver of psychological change. Hill et al. describe it as a "conscious meaning shift involving new connections" [1, p. 442], noting that early insights often begin as simple realisations before deepening into more complex higher-dimensional forms (such as emotional understanding, cognitive restructuring, and reframing past experiences). They emphasise that the concept lacks a universally accepted definition and varies across therapeutic approaches. This ambiguity positions insight as both central and flexible, making it suitable for computational modelling when carefully scoped.

For the purposes of this project, these theoretical limitations (lack of consensus, variation across schools, and multi-dimensionality) clarify which components can be practically extracted from text. PsychExtract therefore focuses exclusively on extracting early-stage insight, which is typically expressed through observable language patterns such as emotional descriptions, reflective statements, and self-evaluative comments. These early meaning shifts are the aspects of insight most consistently expressed through text and most amenable to natural language analysis.

This contextualises why insight extraction matters. If early insight influences therapeutic progress, then summarising indicators of such insight from client reflections can help therapists track meaningful shifts over time. This first section of literature review establishes the theoretical motivation for the system, that is, insight is important, language is one of the primary ways it appears, and early insight is computationally detectable.

Emotion as a Core Component of Therapeutic Change

Greenberg and Pascual-Leone argue that emotional processing is a primary driver of change across therapeutic modalities [2]. They outline a process involving emotional awareness, regulation, transformation, and meaning-making. This typically requires clients to articulate internal emotional experiences. While therapists are trained to detect emotional cues, much emotional content is communicated implicitly through language, making consistency and standardization difficult in practice.

This challenge motivates PsychExtract. If written reflections, such as journals, homework tasks, and progress updates, contain emotional signals that contribute to therapeutic insight, then automated extraction of emotional and linguistic patterns can support therapists by ensuring these signals are made visible and consistently interpreted. The aim is not to replace therapist judgement. Rather, PsychExtract produces structured summaries of emotional and cognitive patterns derived from client text. These summaries can draw attention to potential therapeutic themes without making clinical claims, maintaining alignment with ethical guidance in the field. This naturally leads to the next question of whether artificial intelligence is a suitable tool for supporting therapists in recognising these linguistic signals.

AI Support in Psychotherapy: Lessons from Existing Systems

The use of artificial intelligence in mental-health contexts is not new. DeVault et al. introduced SimSensei [3], a virtual interviewer designed to detect psychological distress from verbal and

nonverbal behaviour. Their work demonstrates several key findings relevant to PsychExtract. Notably, people often disclose more openly when interacting with automated systems, and even simple computational methods can highlight meaningful psychological cues (such as sentiment shifts or linguistic markers of distress).

SimSensei's limitations are equally informative. Because it operates in real-time conversation, it must use extremely cautious and overly simplistic natural language models to avoid unsafe or inappropriate responses. As a result, the system relies on basic language processing.

PsychExtract diverges from this setting in two important ways. First, it is non-conversational. Users provide reflective text, and the system produces an analysis, not an ongoing dialogue. Second, it does not operate in real-time. These affordances allow PsychExtract to employ more advanced language-processing techniques safely, such as transformer-based architectures like BERT, because there is no risk of generating incorrect or harmful conversational replies.

This section therefore establishes why artificial intelligence is appropriate for insight extraction. Existing work shows that AI can highlight clinically relevant linguistic cues, and PsychExtract extends this by applying stronger models in a safer, offline workflow. This bridges into the next section by motivating how AI can be used. This is through specific natural language processing methods tailored to the linguistic components that make up early insight.

Natural Language Processing (NLP) Methods for Emotion, Insight, and Cognitive Pattern Extraction

Before examining technical methods, it is important to clarify terminology for non-specialist readers. Natural language processing (NLP) refers to computational techniques for analysing or generating human language. Modern NLP often uses transformer-based models, which are deep learning architectures capable of understanding words in context rather than in isolation. These models outperform traditional techniques in tasks involving emotion recognition, topic inference, and meaning extraction, all of which are relevant to early insight.

This section details the three components of insight that PsychExtract identifies through NLP: Emotion expression, cognitive themes and reflective topics, and linguistic patterns associated with meaning-making.

By connecting these components to the earlier theory section, PsychExtract grounds its extraction pipeline directly in the psychological mechanisms of insight.

Fine-Grained Emotion Classification (GoEmotions)

Demszky et al. introduce GoEmotions, a dataset of 58,000 Reddit comments labelled with 27 fine-grained emotion categories excluding a neutral class [4]. Their findings show that transformer-based models such as BERT significantly outperform traditional machine learning approaches for understanding emotional nuance, especially because emotions often overlap and require contextual interpretation.

GoEmotions is valuable as a baseline for PsychExtract, but it has limitations. It contains short social-media comments rather than long reflective writing, and deeper therapeutic emotions (such as, grief processing, self-evaluation, growth-related fear) are underrepresented. To address this, PsychExtract uses GoEmotions models for initial benchmarking but extends beyond the dataset by incorporating long-form reflective text. This is in the form of available corpora (such as r/offmychest) or carefully

synthesised paragraphs. This supports the system's goal of aligning emotion extraction with therapeutic contexts.

By grounding the emotional component of insight in this literature, PsychExtract builds directly on empirical evidence that transformer-based models are the strongest choice for contextual emotion detection. This sets the foundation for the next analytic component of understanding cognitive themes.

Cognitive Theme Extraction and Topic Representations (KeyBERT & BERTopic)

Cognitive themes represent the content of what clients reflect on. This is the issues, topics, meanings, and internal processes they describe. To extract these elements, PsychExtract evaluates two widely used NLP tools.

KeyBERT identifies keywords using semantic similarity between the text and candidate n-grams (varied word length groupings) [5]. Because it relies on Sentence-BERT embeddings, it captures meaning beyond simple word counts and is transparent enough to be interpretable by therapists. This makes KeyBERT a suitable, explainable baseline for cognitive theme extraction.

However, KeyBERT provides surface-level patterns and cannot capture broader shifts in meaning across a document. To complement this, PsychExtract includes a comparison with BERTopic, which identifies themes using clustering and class-based term frequency [6]. While more complex, BERTopic can represent broader reflective patterns that align with cognitive restructuring processes described in psychotherapy literature.

This therefore connects the "what" of insight (cognitive content) with the "how" of extraction (topic modelling methods), completing the second component of insight analysis. The next step is to capture linguistic patterns associated with meaning-making.

Linguistic Pattern Analysis (LIWC)

The Linguistic Inquiry and Word Count (LIWC) framework categorises words into psychological dimensions such as cognitive processes, emotional tone, pronoun use, and insight-related terms [7]. Decades of studies demonstrate that these categories reflect internal cognitive states and are particularly relevant for detecting reflective thinking.

PsychExtract draws specifically on the cognitive mechanisms category. Words like "think," "realise," or "because" often signal reflective insight processes. However, LIWC is limited by its dictionary-based approach. It counts words without understanding context. This means it cannot distinguish between "I think" used casually versus reflectively.

To address this limitation, PsychExtract uses LIWC only for interpretability and theoretical grounding, while relying on contextual models (such as transformer-based NLP architectures) for the actual extraction pipeline. This hybrid approach supports interpretability without sacrificing nuance.

Having established how the system extracts early insight from text (emotion, cognitive themes, and linguistic patterns), the final section explains what text is fed into the system and how the results are returned to the user. This completes the OCR-NLP-TTS pipeline.

Input and Output Processing

OCR Requirements in Mental-Health Tools

In therapeutic settings, clients often maintain handwritten journals or written reflections. To analyse such inputs computationally, they must first be digitised using Optical Character Recognition (OCR), a technology that converts images of text into machine-readable characters.

Smith provides a foundational overview of Tesseract, one of the most widely used open-source OCR engines [8]. Tesseract uses a multi-stage pipeline involving line detection, character segmentation, and language modelling to recognise text, even from noisy or imperfect inputs. However, Smith identifies two key limitations. For one, handwriting varies significantly between users, and for two, errors introduced by OCR can propagate into downstream NLP tasks, affecting emotion classification or topic modelling.

PsychExtract incorporates these findings by explicitly evaluating how OCR performance impacts the accuracy of insight-related NLP outputs. This extends prior OCR literature by shifting the focus from character-level accuracy to its influence on psychological inference quality. This is a critical factor in real-world mental-health tooling.

Output Processing: Text-to-Speech Synthesis of NLP Summaries

Once textual insight has been extracted, PsychExtract produces an accessible output for users. Shen et al. introduced Tacotron 2, a leading text-to-speech (TTS) model capable of generating highly natural-sounding audio using a sequence-to-sequence architecture and a neural vocoder [9]. Their work demonstrated that TTS systems can reliably convert text into expressive speech.

In PsychExtract, TTS is used not for interaction but as an accessibility feature. The system reads out the generated summaries, emotional indicators, and cognitive themes for users who prefer auditory feedback or have reading difficulties. This completes the pipeline by providing an intuitive and inclusive output format.

Together, the OCR-NLP-TTS structure forms the full workflow. Handwritten or typed text is digitised; emotional, cognitive, and linguistic markers of early insight are extracted; and results are returned as both text and spoken summaries.

References

[1] Clara E. Hill, Louis G. Castonguay, Lynne Angus, *et al*. 2007. Insight in psychotherapy: Definitions, processes, consequences, and research directions. In *Insight in psychotherapy*. Louis G. Castonguay and Clara E. Hill (Eds.), 441–454. American Psychological Association. https://doi.org/10.1037/11532-021

[2] Leslie S. Greenberg and Antonio Pascual-Leone. 2006. Emotion in psychotherapy: A practice-friendly research review. *J. Clin. Psychol.* 62, 5 (2007), 611–630. https://doi.org/10.1002/jclp.20252

[3] David DeVault, Ron Artstein, Grace Benn, *et al*. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of AAMAS '14*, May, 2014, Paris, France, 1061–1068. https://dl.acm.org/doi/10.5555/2615731.2617415

[4] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, *et al*. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of ACL*, July, 2020, Online. 4040–4054. https://doi.org/10.18653/v1/2020.acl-main.372

[5] Martin Grootendorst. 2020. KeyBERT: Minimal Keyword Extraction with BERT. Retrieved from https://github.com/MaartenGr/KeyBERT, archived at https://web.archive.org/web/20251123033346/https://github.com/MaartenGr/KeyBERT

[6] Martin Grootendorst. 2022. BERTopic: Neural Topic Modelling with Transformers and class-based TF-IDF. *Manuscript submitted for review*. https://doi.org/10.48550/arXiv.2203.05794 archived at https://web.archive.org/web/20250808005533/https://arxiv.org/abs/2203.05794

[7] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, *et al*. 2015. The Development and Psychometric Properties of LIWC2015. *University of Texas at Austin*. https://doi.org/10.13140/RG.2.2.23890.43205

[8] Ray W. Smith. 2007. An Overview of the Tesseract OCR Engine. In *Proceedings of ICDAR '07 Vol. 2*, Sep, 2007, USA, 629–633. https://doi.org/10.1109/ICDAR.2007.4376991

[9] Jonathan Shen, Ruoming Pang, Ron J. Weiss, *et al*. 2018. Natural TTS Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions. In *Proceedings of ICASSP '18*, Apr, 2018, 4779–4783. https://doi.org/10.1109/ICASSP.2018.8461368