PsychExtract: Draft Report

COURSE
CM3070 Final Project
Computer Science
University of London

X words

ABSTRACT

Psychotherapeutic practice increasingly relies on reflective analysis of client narratives, yet manual extraction of emotional and cognitive patterns from written material remains time-intensive and subjective. PsychExtract investigates the feasibility of an AI-assisted pipeline designed to support psychotherapeutic reflection by automatically extracting emotional signals, insight-related themes, and interpretable summaries from client-generated text. The system integrates optical character recognition (OCR) for handwritten and scanned journal entries, natural language processing (NLP) techniques for fine-grained emotion classification and thematic extraction, and an output layer that provides both textual and text-to-speech (TTS) audio summaries. This preliminary report grounds the system design in established psychological theory, drawing on research into insight as a mechanism of therapeutic change and the role of emotion in cognitive restructuring. Existing AI-supported mental-health tools are reviewed to identify methodological and ethical constraints, informing the system's design principles and evaluation strategy. The implemented architecture combines OCR preprocessing, transformer-based emotion classification inspired by the GoEmotions framework, keyword and topic extraction methods for interpretability, and text-to-speech synthesis to improve accessibility. A functional prototype is developed with particular emphasis on OCR feasibility, as transcription quality directly affects all downstream analyses. The prototype is evaluated using quantitative metrics and qualitative error analysis to assess recognition performance, preprocessing effects, and system robustness. Results demonstrate that while OCR performance varies across handwriting styles, targeted preprocessing and error analysis substantially improve downstream interpretability. The findings highlight both the promise and limitations of deploying NLP-driven analysis in sensitive mental-health contexts, and motivate future work on multimodal evaluation, user-centred validation, and clinical collaboration.

CONTENTS

Fix:

Some repetition between:

- Introduction ↔ Design overview
- Literature synthesis ↔ Design principles

This is not wrong, but it contributes to word pressure.

Design section exceeds limit

- Compress the opening overview (it overlaps with Intro)
- Trim repeated ethical constraints already established earlier
- Slightly condense Domain- and User-Framing (keep assumptions, cut re-explanation)

You don't need to lose content, just tighten.

Right now, your Implementation section still reads more like a design-extension than an implementation report.

- Concrete algorithm descriptions
- Actual data flow
- Real failure cases
- Screenshots / outputs tied to implemented code

INTRODUCTION (984/1000 WORDS)

PROJECT OVERVIEW AND MOTIVATION

Psychological reflection and therapeutic writing play a central role in supporting emotional awareness, insight formation, and personal growth. Across both formal psychotherapy and informal self-reflective practices, individuals regularly produce handwritten or semi-structured texts such as journals, therapy notes, worksheets, and reflective exercises (Clara E. Hill et al, 2007; Leslie S. Greenberg and Antonio Pascual-Leone, 2006). These artefacts often contain rich emotional cues, recurring linguistic patterns, and implicit indicators of cognitive framing. Despite their potential value, however, such materials are rarely analysed in a systematic or structured way. Insight extraction remains largely manual, subjective, and dependent on time-intensive human interpretation (Theresa A. Koleck et al, 2019; Sankha S. Mukherjee et al, 2020; Rosanne J. Turner et al, 2022).

This challenge is particularly evident in the case of handwritten material. While digital mental health tools and text-based analytics have expanded rapidly (Becky Inkster et al, 2020; John Torous et al, 2018), a substantial proportion of reflective writing still occurs on paper, creating a disconnect between expressive practices and computational support. Manual review does not scale well, is prone to inconsistency, and may fail to surface longer-term patterns across documents, contributing to cognitive and administrative burden for practitioners and users alike (Tait D. Shanafelt et al, 2016). At the same time, fully automated interpretation of psychological content raises ethical concerns, particularly when systems move beyond assistance into diagnosis or decision-making (Luxton, David D, 2014; Miner, Adam S et al, 2016; Saif M. Mohammad, 2022; MHRA, 2025). These tensions motivate the need for assistive tools that can structure reflective content while preserving transparency and human oversight.

PsychExtract is developed in response to this gap. The project investigates whether handwritten, mental-health-related text can be transformed into structured psychological insight through a modular, interpretable computational pipeline. Rather than attempting to infer mental health conditions or provide therapeutic recommendations, the system focuses on extracting signals already present in the text (such as emotional tone, recurring themes, and linguistic framing) and presenting them in a form that supports reflection. This positioning aligns with ethical guidance emphasising assistive, non-diagnostic AI systems in mental health contexts (Luxton, David D, 2014; Saif M. Mohammad, 2022; Maia Jacobs et al., 2021).

PROJECT TEMPLATE

This project follows Template 4.1: Orchestrating AI Models to Achieve a Goal.

Template 4.1 is particularly appropriate for PsychExtract because the project's contribution lies in system-level integration rather than the development of a single novel model. Prior work in clinical and therapeutic NLP demonstrates that meaningful insight often emerges from combining multiple complementary techniques (such as emotion classification, linguistic analysis, and topic modelling) rather than relying on a single predictive model (Theresa A. Koleck et al, 2019; Sankha S. Mukherjee et al, 2020; Becky Inkster et al, 2020; Finale Doshi-Velez and Been Kim, 2017). Each stage of the pipeline (optical character recognition, emotion classification, linguistic analysis, semantic summarisation, and optional accessibility output) offers multiple viable modelling approaches. The template provides a structured framework for selecting, integrating, and evaluating these components within a unified system.

This orchestration-focused approach aligns with research advocating transparent, modular AI systems in sensitive domains, where trust and explainability are as important as technical performance (Maia Jacobs et al., 2021; Finale Doshi-Velez and Been Kim, 2017).

PROJECT AIMS AND OBJECTIVES

The overarching aim of PsychExtract is to support reflective understanding of handwritten text through structured and interpretable computational analysis. The project is explicitly non-clinical in nature. It does not seek to diagnose mental health conditions, predict outcomes, or offer therapeutic advice, consistent with regulatory and ethical guidance for AI systems that operate outside formal medical device classification (Luxton, David D, 2014; Saif M. Mohammad, 2022; MHRA, 2025).

To achieve this aim, the project designs a modular processing pipeline that converts handwritten reflective text into meaningful digital representations. Optical character recognition (OCR) forms the foundational stage of this pipeline, as inaccuracies at this level directly affect all subsequent processing. Following digitisation, the system applies emotion classification using transformer-based models selected for their balance between performance and interpretability (Dorottya Demszky et al, 2020; Victor Sanh et al, 2019; Liu Yinhan et al, 2019; Jacob Devlin et al, 2019). Linguistic pattern extraction is then employed to identify recurring keywords, phrasing, and framing strategies that may signal cognitive or emotional emphasis (Maarten Grootendorst, 2025; Maarten Grootendorst, 2022; James W. Pennebaker et al, 2015). These signals are integrated into concise semantic summaries that prioritise clarity and human readability over abstraction.

In addition to analytical objectives, the project also aims to explore accessibility and user experience considerations. This includes the optional integration of text-to-speech (TTS) functionality. Prior work demonstrates that TTS can support accessibility and reduce reading fatigue, particularly for users with learning differences or cognitive load constraints (Jonathan Shen et al, 2018; Mary Cece Young et al, 2018).

TARGET DOMAIN AND USERS

PsychExtract operates within the broader domain of mental health support, self-reflection, and therapeutic-adjacent technologies. Its focus is not on clinical intervention but on assisting reflective practices where emotional awareness and insight formation are valuable (Clara E. Hill et al, 2007; Leslie S. Greenberg and Antonio Pascual-Leone, 2006). The system is intended for use in contexts where users benefit from structured reflection but where automated clinical judgement would be inappropriate or ethically problematic (Luxton, David D, 2014; Saif M. Mohammad, 2022).

The primary target users include students engaging in reflective writing as part of their academic or personal development, individuals who journal as a form of emotional processing, and therapy-adjacent users who reflect on experiences between sessions or prepare written material for discussion. Similar user groups have been identified in prior research on digital mental health tools and reflective technologies (Becky Inkster et al, 2020; John Torous et al, 2018; Dror Ben-Zeev et al, 2013). In all cases, the system is positioned as an assistive tool rather than an authority. It structures and summarises content supplied by the user without interpreting intent, assigning diagnoses, or offering recommendations.

Maintaining clear boundaries around system capability is a central design principle of the project. PsychExtract does not claim psychological understanding or therapeutic expertise. Instead, it supports users in engaging more effectively with their own writing, reinforcing reflection rather than replacing it.

Contributions and Originality

The originality of PsychExtract lies in its integration of multiple AI and NLP techniques into a coherent, reflection-oriented pipeline tailored to handwritten text. While individual components such as OCR, emotion classification, and topic modelling are well-established (Dorottya Demszky et al, 2020; Maarten Grootendorst, 2022; Ray W. Smith, 2007), their combination within a non-clinical, interpretability-focused framework represents a novel application aligned with ethical AI design principles (Saif M. Mohammad, 2022; Maia Jacobs et al., 2021; Finale Doshi-Velez and Been Kim, 2017).

The project contributes a system that bridges the gap between expressive handwritten practices and computational analysis, demonstrating how structured insight extraction can be achieved without overstepping ethical boundaries. Its emphasis on transparency, modularity, and accessibility distinguishes it from systems that prioritise prediction or automation alone. By foregrounding interpretability and future extensibility, particularly through planned interface development and accessibility features such as text-to-speech, PsychExtract highlights how AI systems can be responsibly designed to support sensitive human activities.

Literature Review (1719/2500)

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" (Clara E. Hill et al, 2007, 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 (Leslie S. Greenberg and Antonio Pascual-Leone, 2006). 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 (David DeVault et al. 2014), 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).

Figure 1. Ellie, the virtual human interviewer used in the SimSensei Kiosk system (David DeVault et al, 2014)

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.

NLP Methods for Emotion, Insight, and Cognitive Pattern Extraction

Before examining technical methods, it is important to clarify terminology for non-specialist readers. 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 (Dorottya Demszky et al, 2020), visualized in Figure 2. 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.

Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

Figure 2. GoEmotions emotion taxonomy comprising 28 fine-grained emotion categories, including a neutral class (Dorottya Demszky et al, 2020)

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, which are designed to preserve emotional coherence without introducing clinical claims. 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 and 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 cosine semantic similarity between the text and candidate n-grams (varied word length groupings) (Maarten Grootendorst, 2025), this is visualized in Figure 3. 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.

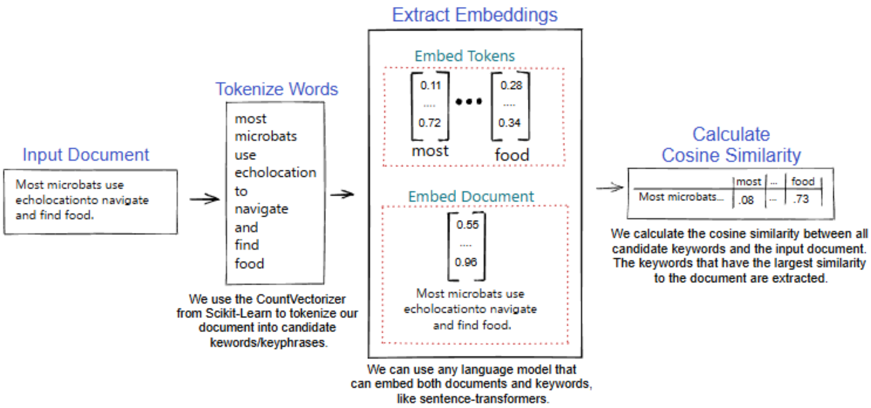

Figure 3. KeyBERT keyword extraction pipeline (Maarten Grootendorst, 2025)

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 (Maarten Grootendorst, 2022), this is illustrated in Figure 4. While more complex, BERTopic can represent broader reflective patterns that align with cognitive restructuring processes described in psychotherapy literature.

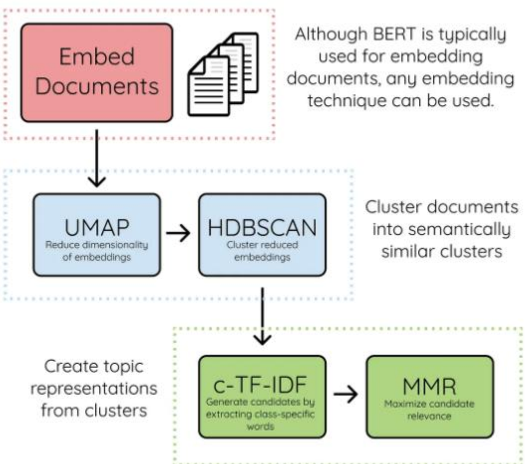

Figure 4. BERTopic topic modelling pipeline (Maarten Grootendorst, 2022)

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 (James W. Pennebaker et al, 2015). 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 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 (Ray W. Smith, 2007). Tesseract uses a multi-stage pipeline (concisely laid out in Figure 5) 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.

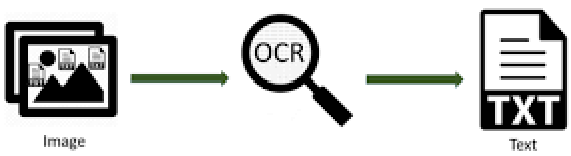

Figure 5. OCR process

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 TTS model capable of generating highly natural-sounding audio using a sequence-to-sequence architecture and a neural vocoder (Jonathan Shen et al. 2018). Their work demonstrated that TTS systems can reliably convert text into expressive speech. The pipeline is summarized in Figure 6.

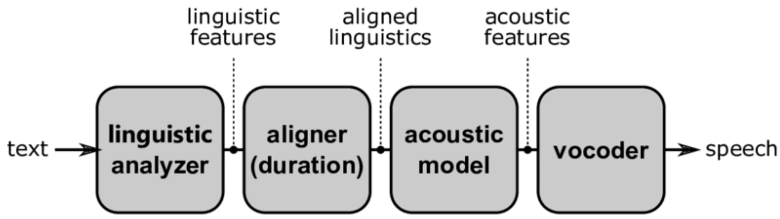

Figure 6. Conventional TTS pipeline representation

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.

CRITICAL SYNTHESIS: OPERATIONALIZING INSIGHT RESPONSIBLY

The reviewed literature demonstrates that while NLP techniques can reliably extract linguistic signals associated with emotion, reflection, and thematic content, each approach is subject to important limitations. Emotion classification relies on probabilistic labels that simplify context-dependent and culturally mediated experiences; keyword extraction and topic modelling prioritise semantic salience over psychological coherence; and dictionary-based linguistic tools lack contextual sensitivity. Furthermore, many existing AI systems in mental-health contexts are constrained by real-time interaction requirements, limiting the complexity of language processing that can be deployed safely.

These limitations directly motivate the integrated, interpretability-focused design of PsychExtract. Rather than attempting to infer insight or psychological state directly, the system positions itself as an assistive analytic tool that surfaces multiple complementary indicators of early insight from reflective text. By combining emotion signals, cognitive themes, and linguistic markers within a non-conversational, offline pipeline, PsychExtract prioritises transparency, theoretical grounding, and human oversight. This modular, hybrid architecture reflects a deliberate shift away from diagnostic or interventionist claims toward a supportive, human-in-the-loop role that aligns with ethical guidance for AI use in therapeutic contexts (Luxton, David D, 2014; Miner, Adam S et al, 2016.

DESIGN (2172/2000 WORDS)

REDUCE WORDS

OVERVIEW

PsychExtract is designed as an interpretability-first, non-clinical analysis system intended to support structured reflection on emotionally expressive text. The core objective of the system is not to provide psychological judgement or intervention, but to assist users in identifying emotional patterns, salient themes, and linguistically meaningful signals within their own writing. This positioning aligns with established work emphasising insight formation as a central mechanism of psychological change, while cautioning against automated clinical interpretation (Hill et al., 2007; Greenberg and Pascual-Leone, 2006; Luxton, 2014).

At a system level, PsychExtract aims to transform unstructured textual material (such as journals, reflective essays, or therapy-adjacent notes) into structured, inspectable outputs that can prompt reflection without asserting authority. Rather than collapsing analysis into a single opaque prediction, the system exposes intermediate representations at each stage of processing. This design choice reflects both the subjective nature of emotional language and consistent findings that transparent, inspectable systems are more likely to be trusted and appropriately relied upon in sensitive mental-health-adjacent contexts (Doshi-Velez and Kim, 2017; Jacobs et al., 2021).

Several constraints are deliberately embedded in the design. First, PsychExtract is explicitly non-clinical. It does not infer diagnoses, assess risk, or generate therapeutic recommendations. Outputs are framed descriptively and probabilistically, reflecting patterns in language rather than claims about the user's mental state, consistent with ethical guidance for affective computing and mental health technologies (Luxton, 2014; Mohammad, 2022). Second, the system is privacy-aware by design. Processing is intended to occur locally where possible, relying on pre-trained models rather than cloud-based services that require data transmission. Persistent storage of user text is avoided, and intermediate artefacts are treated as transient analysis objects. Third, the system prioritises interpretability over predictive optimisation. While transformer-based models are employed, they are paired with explicit linguistic explanation layers and conservative output framing, favouring clarity and traceability over marginal gains in classification accuracy.

Together, these goals and constraints define the design space within which PsychExtract operates: assistive rather than authoritative, transparent rather than optimised, and ethically bounded rather than expansive.

DESIGN PRINCIPLES

The design of PsychExtract is guided by four core principles: transparency, modularity, error tolerance, and user trust. These principles are embedded not only in interface decisions but in the underlying system architecture and data flow.

Transparency is prioritised through the exposure of intermediate representations and the avoidance of opaque, end-to-end inference. Users are able to inspect OCR output, view emotion probabilities rather than categorical labels, and see which linguistic features contribute to explanatory summaries. This approach reflects the distinction between interpretability and raw model introspection, favouring explanations that are intelligible and meaningful to non-expert users (Doshi-Velez and Kim, 2017; Mohammad, 2022).

PsychExtract is structured as a collection of loosely coupled components, each responsible for a specific transformation of the data, encouraging a modular design. This modularity allows alternative OCR engines, language models, or keyword extraction techniques to be substituted without rearchitecting the system as a whole. Such flexibility supports iterative development, comparative evaluation, and future extension, and mirrors best practices observed in clinician-facing NLP pipelines (Turner et al., 2022).

Given the inherently noisy nature of reflective writing and upstream processes such as OCR, the system is designed to tolerate and contain errors rather than obscure them. User verification checkpoints, parallel analytical pathways, and probabilistic output framing all contribute to limiting the downstream impact of misrecognition or misclassification. By avoiding hard thresholds and definitive claims, the system reduces the risk of misleading or overconfident outputs in the presence of uncertainty.

Trust is treated as a design outcome rather than a user assumption. By maintaining clear non-clinical boundaries, minimising data retention, and framing outputs as descriptive rather than prescriptive, PsychExtract aims to encourage appropriate reliance. Prior work emphasises that trust in AI-supported mental health tools is fostered when system limitations are legible and role boundaries are respected (Luxton, 2014; Jacobs et al., 2021; Torous et al., 2018).

These principles serve as the conceptual constraints guiding the subsequent domain framing, component selection, and architectural decisions

DOMAIN- AND USER-FRAMING IN DESIGN

PsychExtract operates within the broader domain of mental health support, self-reflection, and therapeutic-adjacent technologies. Its focus is not on clinical intervention but on assisting reflective practices where emotional awareness and insight formation are valuable. Prior research highlights the potential of structured language analysis to support psychological understanding while simultaneously emphasising the risks of over-automation and misinterpretation in mental health contexts (Inkster et al., 2020; Torous et al., 2018; Mohammad, 2022).

The system is intended for users who already engage in reflective writing but may benefit from additional structure and pattern visibility. These users include students completing reflective assignments, individuals who journal for emotional processing, and therapy-adjacent users who write between sessions or prepare material for discussion. Across these groups, the system assumes voluntary engagement and user ownership of interpretation. PsychExtract structures and summarises content supplied by the user without interpreting intent, assigning diagnoses, or offering recommendations, reinforcing its role as an assistive tool rather than an authority.

Several explicit assumptions and exclusions guide the design. Users are assumed to be reflective rather than crisis-seeking, and the system does not attempt to detect self-harm, suicidality, or acute distress. It does not provide real-time intervention, escalation pathways, or safeguarding mechanisms. While clinicians and researchers may find the system legible and methodologically informative, they are not positioned as primary end-users in this iteration. Maintaining these boundaries is a deliberate design decision intended to preserve trust, reduce ethical risk, and avoid role confusion, consistent with guidance on responsible AI use in mental health settings (Luxton, 2014; MHRA, 2025).

SYSTEM COMPONENTS

PsychExtract is organised as a modular pipeline composed of interoperable components (OCR, NLP with an interpretability layer, and TTS), each responsible for a specific transformation of the data. This architectural choice reflects findings from prior work showing that clinicians and users prefer systems that expose intermediate reasoning steps and allow scrutiny of automated outputs

(Jacobs et al., 2021; Turner et al., 2022). It also supports comparative evaluation by allowing alternative models and techniques to be substituted without restructuring the entire system.

OCR

The OCR component converts scanned documents, photographs, or PDFs containing handwritten or typed text into machine-readable form. Two open-source OCR engines are considered: Tesseract and EasyOCR. Tesseract is a lightweight, widely used engine with strong performance on clean, printed text (Smith, 2007; Patel et al., 2012), while EasyOCR employs deep learning architectures that are more robust to noisy inputs and variable handwriting styles (JaidedAI, 2024). Comparing these engines allows assessment of robustness across the heterogeneous document quality typical of reflective writing.

The input to this component consists of image-based documents, while the output is plain text. OCR output is explicitly treated as provisional and is surfaced to the user for verification and correction before downstream processing. Integration at this stage is intentionally partial: when users provide typed text directly, the OCR component can be bypassed entirely, reinforcing flexibility in input modality and reducing unnecessary error propagation.

Emotion Classification

Emotion classification is central to the system's analytical goals, as emotional awareness and differentiation are closely linked to insight formation in psychotherapy and reflective practice (Greenberg and Pascual-Leone, 2006; Hill et al., 2007). PsychExtract employs pre-trained transformer-based models fine-tuned for multi-label emotion classification using the GoEmotions dataset (Demszky et al., 2020).

Two models are explicitly considered: DistilBERT and RoBERTa. DistilBERT offers a compressed architecture that retains much of BERT's representational capacity while reducing computational cost (Sanh et al., 2019), making it suitable for resource-constrained environments. RoBERTa, by contrast, benefits from optimised pretraining strategies and typically achieves higher classification performance at the cost of increased computational demand (Liu et al., 2019). Comparing these models supports an explicit trade-off analysis between performance and feasibility.

The input to this component is cleaned, user-verified text, and the output consists of probabilistic multi-label emotion predictions. This is demonstrated in Figure 7. Outputs are framed as signals detected in language rather than definitive emotional states, reflecting both dataset limitations and ethical guidance discouraging overconfident affective inference (Mohammad, 2022).

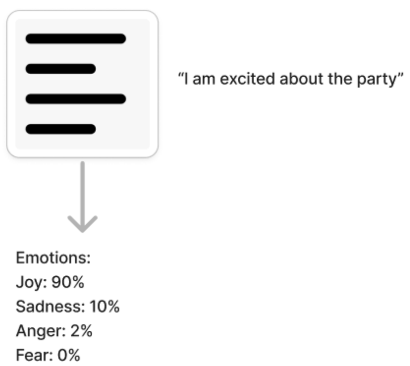

Figure 7: Emotion Classification model input and output visualized

Keyword and Linguistic Pattern Extraction

To contextualise emotion predictions, PsychExtract includes a linguistic feature extraction component that identifies salient words, phrases, or patterns associated with emotional signals. Embedding-based methods such as KeyBERT are employed to capture semantic relevance between text segments and extracted keywords (Grootendorst, 2025). Alternative approaches, including TF–IDF weighting and psychologically motivated lexical markers inspired by LIWC, provide more rule-based and clinically interpretable signals (Pennebaker et al., 2015).

The input to this component consists of text and, optionally, emotion classification outputs, while the output is a set of explanatory linguistic features. This pipeline is demonstrated in Figure 8. This component is only partially coupled to emotion classification: keyword extraction can operate independently, allowing alternative interpretability pathways and supporting error analysis when emotion predictions are uncertain.

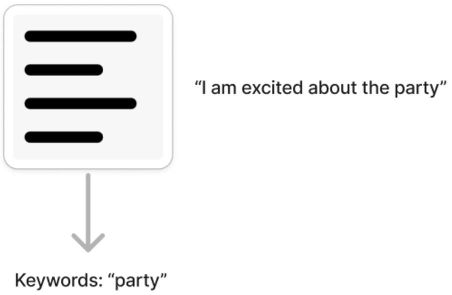

Figure 8: Keyword and Linguistic Pattern Extraction input and output visualized

Interpretability Layer

The interpretability layer synthesises emotion predictions and linguistic features into human-readable explanatory artefacts. Rather than exposing internal model mechanics such as attention weights, which may be misleading or difficult to interpret reliably, this layer focuses on psychologically legible explanations grounded in observable language patterns. This design choice aligns with broader critiques of superficial explainability in machine learning and emphasises intelligibility over technical transparency (Doshi-Velez and Kim, 2017). Importantly, these explanations do not claim to reveal the 'true cause' of emotional expression, but rather offer plausible, inspectable mappings between language use and detected emotional signals

The output of this layer consists of structured explanations that link detected emotional signals to specific language features, reinforcing transparency and supporting appropriate user trust. This interpretability pipeline is visualized in Figure 9.

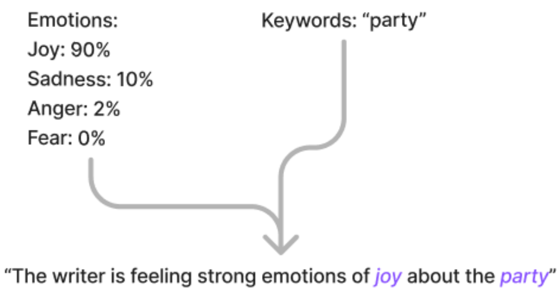

Figure 9: Interpretability layer inputs and output demonstrated

Output Generation and Accessibility

The final stage of the pipeline generates concise textual summaries that highlight dominant emotional themes, recurring linguistic patterns, and notable shifts across the input text. These summaries are designed to prompt reflection rather than provide conclusions. An optional TTS component supports accessibility for users with reading, attentional, or cognitive challenges, consistent with evidence that TTS can enhance comprehension and engagement (Young et al., 2018).

Two TTS approaches are considered: neural synthesis via Coqui TTS, which offers high-quality speech generation (Coqui.ai, 2025), and pyttsx3, which provides a lightweight offline alternative suitable for constrained environments (pyttsx3.readthedocs.io, 2025). The UI at this stage remains deliberately minimal, prioritising clarity, editability, and progressive disclosure of information over visual richness.

Architecture and Data Flow

PsychExtract uses a modular, linear pipeline in which each stage is independently testable and replaceable. This maximises interpretability, facilitates comparative evaluation, and maintains clear traceability across the workflow. Data progresses through five main stages: OCR, emotion classification, linguistic interpretation extraction, summary generation, and optional TTS.

User Flow

Figure 10 illustrates the end-to-end user flow, beginning with raw document ingestion and culminating in a structured insight summary. The pipeline comprises five core computational stages, supported by user-mediated validation and optional output modalities.

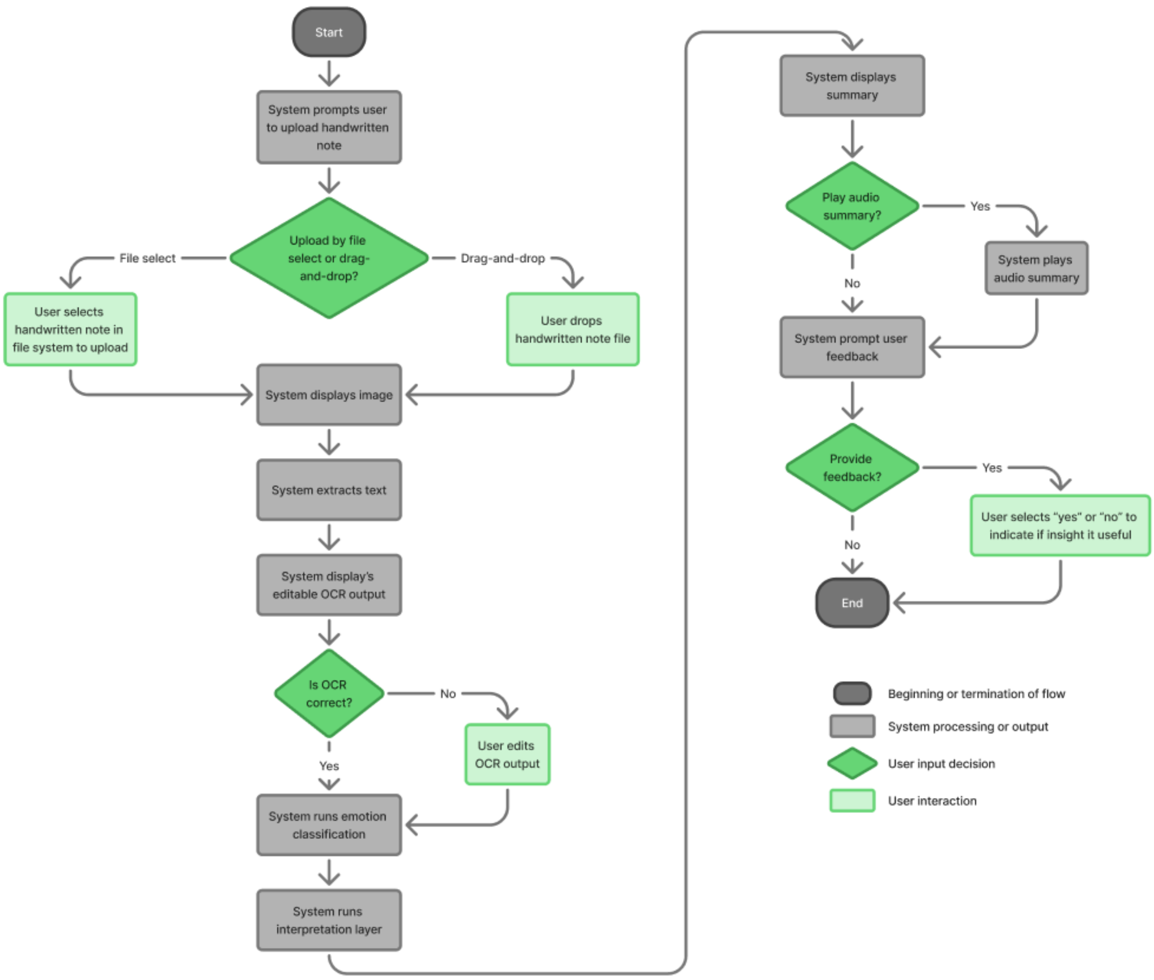

Figure 10. PsychExtract user flow

The following is a summarized breakdown of the steps that the system undertakes:

1. Input Upload: user provides handwritten or scanned document
2. OCR: Tesseract/EasyOCR produces preliminary text
3. Editable Text: user verifies and corrects OCR output.
4. Emotion Classification: DistilBERT/RoBERTa produce multi-label probabilities
5. Interpretability Layer: KeyBERT or linguistic metrics extract explanatory features
6. Summary Generation: concise psychological insight produced
7. Optional TTS: pyttsx3/Coqui converts summary to speech
8. User Feedback Capture: supports iterative refinement and future evaluation

Early Prototype

To explore alternative interaction patterns without committing to full implementation, interfaces are tested in Figma. Figure 11 presents a low-fidelity Streamlit prototype that demonstrates document upload, OCR correction, and insight review workflows. User feedback capture is embedded throughout the interface to support iterative refinement and future evaluation, though this feedback loop is not yet fully integrated into model retraining.

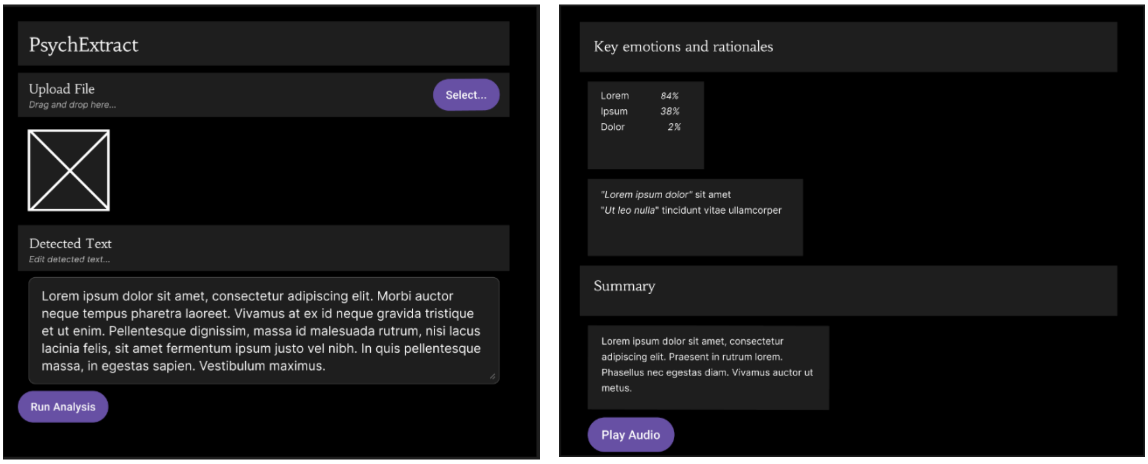

Figure 11. PsychExtract prototype interface

Folder Structure

At the code level, the system is organised into discrete modules reflecting each pipeline stage (Table 1). This folder structure enforces separation of concerns, promotes traceability between components, and supports flexible model substitution during experimentation and evaluation.

Folder	Purpose
ocr/	OCR wrappers for Tesseract and EasyOCR
nlp/	DistilBERT and RoBERTa models
interpret/	KeyBERT and linguistic metrics
summary/	Summarization and TTS components
ui/	Streamlit interface
tests/	User and integration tests

Table 1. PsychExtract folder structure

Feasibility and Constraints

The design of PsychExtract is shaped by several practical constraints relating to technical feasibility, data availability, and project scope.

There are technical constraints present as the system relies on transformer-based language models, which impose computational costs in terms of memory usage and inference time. While models such as DistilBERT mitigate these demands relative to larger architectures, real-time processing on low-resource devices remains constrained. The decision to support local, offline processing further limits the use of large-scale or continuously updated models, but aligns with privacy and ethical goals (Luxton, 2014). Similarly, OCR performance is sensitive to document quality, handwriting variability, and image noise, constraining accuracy in unconstrained real-world inputs (Smith, 2007; JaidedAI, 2024).

Emotion classification is tested on the GoEmotions dataset, which, while large and fine-grained, is derived from curated online text rather than therapeutic or deeply personal writing. This exposes a data limitation where, as a result, emotion predictions may not fully capture the nuance of reflective or autobiographical language. Additionally, emotion labels represent perceived emotional content in text, not ground-truth internal states, reinforcing the need for cautious, probabilistic interpretation (Demszky et al., 2020; Mohammad, 2022).

As an individual academic project with time constraints present, PsychExtract is necessarily scoped to prioritise depth of reasoning over breadth of functionality. Advanced features such as longitudinal user modelling, adaptive feedback, or clinician-facing analytics are intentionally excluded in favour of a robust, interpretable core pipeline. This constraint supports methodological clarity and ethical containment while leaving clear pathways for future work.

IMPLEMENTATION (2000 WORDS)

"At the time of writing, integration is partially implemented…"

IMPLEMENTATION OVERVIEW

- Tech stack
- Modular pipeline philosophy

OCR IMPLEMENTATION

- Tools used
- Preprocessing

EMOTION NLP

- Models used (DistilBERT / RoBERTa)
- Output structure
- Thresholding / aggregation decisions

KEYWORD & LINGUISTIC PATTERN EXTRACTION

- KeyBERT / noun phrase filtering
- Handling redundancy and overlap
- Linguistic templates / framing logic

INTEGRATION STATUS

- What is currently integrated
- What is partially integrated
- What remains conceptual (but designed)

FIGURES

- Example outputs
- Screenshots / tables
- Sample end-to-end flow

Implementation Overview

At the time of writing, PsychExtract is partially implemented as a modular, linear processing pipeline designed to prioritise interpretability, transparency, and component-level traceability. The system architecture reflects an explicit design choice to favour independently testable stages over tightly coupled end-to-end learning, enabling clear inspection of intermediate representations and system behaviour. This design aligns with established principles of interpretable machine learning and human-centred AI systems (Doshi-Velez and Kim, 2017; Jacobs et al., 2021).

The current implementation consists of two realised subsystems: an OCR pipeline for handwritten input and a transformer-based emotion classification module for reflective text. Each subsystem exposes clearly defined input and output interfaces, allowing components to be inspected, substituted, or extended independently without architectural restructuring. This modularity supports both exploratory development and later evaluation-driven refinement.

All components are implemented in Python using widely adopted machine learning and computer vision libraries. Deep learning models are executed using PyTorch and Hugging Face Transformers (Devlin et al., 2019), while OCR experimentation integrates both classical OCR engines and neural-based approaches. Intermediate results are persisted as CSV and JSON artefacts rather than through a unified application interface, reflecting the analytical and research-oriented focus of the current system state.

Figure x: High-level system architecture illustrating the linear data flow from handwritten or textual input through OCR and emotion classification stages.

OCR Implementation

Tools and Model Selection

The OCR subsystem is implemented as a feasibility-oriented pipeline for handwritten text recognition rather than as a production-ready transcription system. Consistent with the scope defined across the report, Tesseract OCR and EasyOCR serve as the primary OCR engines considered for systematic evaluation. Tesseract represents a traditional rule-based OCR baseline (Smith, 2007; Patel et al., 2012), while EasyOCR provides a modern deep learning–based alternative utilising CNN–RNN architectures (JaidedAI, 2024).

During early experimentation, both engines demonstrated substantial difficulty in reliably recognising handwritten reflective text, with character-level accuracy remaining low across samples. In response, additional OCR approaches (PaddleOCR: Cui et al, 2025; TrOCR: Li, M. et al, 2021; and Qwen2.5-VL: Bai et al, 2024) were incorporated on an exploratory basis to assess whether alternative architectural paradigms might better accommodate handwritten input. These models were not adopted as primary evaluation targets but were included to probe the practical limits of handwritten OCR within the project context.

This expanded model set spans classical OCR, specialised handwritten transformer models, and general-purpose vision–language systems, enabling architectural comparison without committing to full downstream integration.

Figure x: Example handwritten input samples used across all OCR engines to ensure consistent comparative input.

Image Preprocessing Pipeline

Prior to OCR, all handwritten images undergo a deterministic preprocessing pipeline intended to improve text visibility while preserving the original spatial structure of handwriting. Images are normalised for orientation using available metadata, converted to grayscale, and contrast-enhanced using Contrast Limited Adaptive Histogram Equalisation (CLAHE) (Zuiderveld, 1994). Adaptive thresholding is then applied to separate foreground text from background noise (Otsu, 1979), followed by optional upscaling to improve character resolution for OCR engines sensitive to input size (Plamondon and Srihari, 2000).

The same preprocessing steps are applied uniformly across all OCR engines to ensure consistency of inputs. Preprocessed images are saved as standalone artefacts, supporting reproducibility and visual inspection of the transformation process.

Figure x: Example handwritten image before and after preprocessing, illustrating contrast enhancement and background suppression applied prior to OCR.

Output Format and Post-processing

Each OCR engine produces a plain-text transcription which is saved as an independent file. For OCR systems that generate structured outputs (such as bounding boxes or conversational responses) only the recognised textual content is retained to maintain consistency across engines.

Lightweight post-processing is applied uniformly to all OCR outputs. This includes case normalisation, punctuation removal, and whitespace standardisation. These steps are applied to reduce superficial variability and to ensure compatibility with downstream natural language processing components. No semantic correction, spell-checking, or content normalisation is performed, preserving fidelity to the OCR-generated text.

Figure x: Example OCR outputs from different engines demonstrating variation in character segmentation and transcription style.

EMOTION NLP IMPLEMENTATION

Model Selection and Rationale

The emotion analysis module employs two pretrained transformer-based multi-label emotion classifiers: DistilBERT and RoBERTa. DistilBERT provides a compact, computationally efficient architecture derived from BERT (Sanh et al., 2019), while RoBERTa offers improved representational capacity through robust pretraining strategies (Liu et al., 2019).

These models were selected to balance efficiency, interpretability, and emotional coverage. DistilBERT aligns with the project's core six-emotion framework (joy, sadness, anger, fear, surprise, love), while RoBERTa supports a broader emotional taxonomy (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust) derived from large-scale emotion-labelled corpora such as GoEmotions (Demszky et al., 2020).

Figure x: Emotion taxonomy comparison

Both models are used strictly in inference-only mode. No fine-tuning is performed within this project, reflecting an emphasis on system integration, transparency, and output behaviour rather than task-specific optimisation.

Figure x: Overview of the emotion classification pipeline, showing tokenisation, model inference, and structured output generation.

Input Processing and Output Structure
Input text is tokenised using the corresponding pretrained tokenizer and truncated to a fixed maximum sequence length. Emotion predictions are generated via sigmoid activation over model logits ,which is done to produce independent probability scores for each emotion label.

Outputs are explicitly structured: for each input text, the system records per-emotion probability scores alongside thresholded binary indicators. All outputs are stored in CSV format, enabling downstream inspection, aggregation, and visualisation without reliance on opaque internal representations.

Figure x: Example emotion prediction table showing per-label probability scores and binary indicators for a single input text.

Thresholding and Aggregation Decisions
Emotion probabilities are converted to binary predictions using a fixed threshold of 0.5. This value is applied consistently across both models to ensure comparability and interpretive stability. No adaptive or class-specific thresholding is applied at this stage.

For RoBERTa outputs, which include a larger emotion set, predictions are mapped onto the project's six core emotion categories. Where multiple labels map to a single category, the maximum confidence score is retained. This approach preserves the strongest emotional signal while avoiding dilution through averaging.

EVALUATION (2500 WORDS)

EVALUATION STRATEGY

- Why mixed evaluation methods are appropriate
- Limitations of large-scale quantitative evaluation

COMPONENT-LEVEL EVALUATION

- OCR accuracy and errors
- Emotion classification behaviour
- Keyword / linguistic extraction quality

USER EVALUATION

- Survey design
- Participant description (n = 2, explicitly stated)
- Qualitative feedback themes

RESULTS

- Tables or excerpts
- Observed strengths
- Observed confusion or limitations

ANALYSIS

- What works well
- Where the system struggles
- Mismatch between intent and perception

LIMITATIONS OF THE EVALUATION

- Sample size
- Prototype fidelity
- Time constraints

OCR

Evaluation and Aggregation Logic

OCR performance is evaluated using multiple complementary metrics:

- **Sequence similarity ratio** using character-level alignment
- **Character-level accuracy**, accounting for insertion and deletion errors
- **Word Error Rate (WER)** using the JiWER framework

Metrics are computed per-sample and aggregated as mean scores across the evaluation set. In addition to scalar metrics, character-level error rates are computed and visualised as a heatmap, highlighting systematic recognition failures for specific characters across models.

This multi-level evaluation strategy avoids reliance on a single performance metric and provides insight into both quantitative accuracy and qualitative error patterns.

Failure Cases and Limitations

Several consistent OCR failure modes were observed across models. Classical OCR systems such as Tesseract exhibited significant degradation on cursive or irregular handwriting, frequently fragmenting words or misclassifying character boundaries. CNN-based systems showed improved robustness but remained sensitive to character overlap and unconventional letter forms.

Transformer-based models demonstrated stronger holistic recognition but introduced different failure modes, including hallucinated punctuation, incorrect word segmentation, and semantic normalisation of text that diverged from the original handwritten input. The vision-language model occasionally produced paraphrased or inferred text rather than strict transcription, despite explicit prompting to avoid correction.

These failure cases highlight a central trade-off in OCR model selection: accuracy versus faithfulness to original text. Given PsychExtract's reliance on downstream emotional and linguistic analysis, transcription errors—particularly negation loss or word substitution—pose a risk of semantic distortion. As a result, the OCR subsystem is currently treated as an optional and carefully constrained input mechanism rather than a fully integrated default component.

EMOTION CLASSIFICATION

Evaluation and Observed Behaviour

Evaluation is conducted using macro- and micro-averaged precision, recall, and F1 scores, alongside per-label metrics. Results demonstrate that both models exhibit uneven performance across emotion categories, with higher reliability for high-frequency emotions such as joy and sadness, and reduced recall for more context-dependent emotions such as surprise.

Importantly, the system does not attempt to resolve conflicting emotional signals into a single dominant label. Multi-label outputs are retained intentionally, reflecting the complexity and ambiguity inherent in emotionally expressive text.

Failure Cases and Interpretive Risks

Several failure patterns were identified. Both models occasionally assign moderate confidence to semantically adjacent emotions, leading to broad emotional profiles rather than sharply differentiated outputs. Sarcasm, implicit emotional cues, and context-dependent sentiment shifts remain challenging, particularly for short or ambiguous texts.

Additionally, model outputs may reflect biases introduced by training data domains, particularly for RoBERTa's social media–oriented corpus. These limitations reinforce the system's positioning as an assistive analytical tool rather than an authoritative emotional assessor.

CONCLUSION AND FUTURE WORK (1000 WORDS)

Strong, reflective, and grounded

PROJECT SUMMARY

- What was built
- What was demonstrated

REFLECTION ON OBJECTIVES

- Which objectives were met
- Which were partially met

FUTURE WORK

- Full integration
- Expanded user studies
- Personalisation

- Ethical extensions

Broader Implications

- Role of NLP in reflective practice
- Interpretability-first mental health tools

References

SORT OUT ALPHABETICALLY
SORT OUT FORMAT

Clara E. Hill, *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

Leslie S. Greenberg and Antonio Pascual-Leone. 2006. Emotion in psychotherapy: A practice-friendly research review. *Journal of Clinical Psychology* 62, 5 (2007), 611–630. https://doi.org/10.1002/jclp.20252

Theresa A. Koleck, *et al.* 2019. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *Journal of the American Medical Informatics Association* 24, 4 (Apr. 2019), 364-379. https://doi.org/10.1093/jamia/ocy173

Sankha S. Mukherjee, *et al.* 2020. Natural language processing-based quantification of the mental state of psychiatric patients. *Computational Psychiatry* 5 (Dec. 2020), 76-106. https://doi.org/10.1162/cpsy_a_00030

Rosanne J. Turner, *et al.* 2022. Information extraction from free text for aiding transdiagnostic psychiatry: constructing NLP pipelines tailored to clinicians' needs. *BMC Psychiatry* 22, Article 407 (Jun. 2022), 11 pages. https://doi.org/10.1186/s12888-022-04058-z

Becky Inkster, *et al.* 2020. Digital Health Management During and Beyond the COVID-19 Pandemic: Opportunities, Barriers, and Recommendations. *JMIR mental health* 7, 7, Article e19246 (Jul. 2020), 5 pages. https://doi.org/10.2196/19246

John Torous, *et al.* 2018. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evidence-based mental health* 21, 3 (Aug. 2018), 116-119. https://doi.org/10.1136/eb-2018-102891

Tait D. Shanafelt, *et al.* 2016. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *In Mayo Clinic Proceedings* 91, 7 (Jul. 2016), 836-848. https://doi.org/10.1016/j.mayocp.2016.05.007

Luxton, David D. "Recommendations for the ethical use and design of artificial intelligent care providers." *Artificial intelligence in medicine* vol. 62,1 (2014): 1-10. https://doi:10.1016/j.artmed.2014.06.004

Miner, Adam S et al. "Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health." *JAMA internal medicine* vol. 176,5 (2016): 619-25. https://doi:10.1001/jamainternmed.2016.0400

Saif M. Mohammad. 2022. Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis. *Computational Linguistics* 48, 2 (Jun. 2022), 239-278. https://doi.org/10.1162/coli_a_00433

Medicines and Healthcare Products Regulatory Agency (MHRA). 2025. Software and artificial intelligence (AI) as a medical device. *gov.uk*. Retrieved Nov 27. 2025 from https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device, *archived at* [https://web.archive.org/web/20250815022341/https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device]

Maia Jacobs, *et al.* 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* Association for Computing Machinery 659, 1-14. https://doi.org/10.1145/3411764.3445385

Finale Doshi-Velez and Been Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint.* https://doi.org/10.48550/arxiv.1702.08608

Dorottya Demszky, *et al.* 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of ACL* (Jul. 2020), 4040–4054. https://aclanthology.org/2020.acl-main.3t2/

David DeVault, *et al.* 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of AAMAS '14* (May. 2014), 1061–1068. https://dl.acm.org/doi/10.5555/2615731.2617415

Maarten Grootendorst. 2025. KeyBERT. *GitHub*. Retrieved Nov. 26 2025 from https://github.com/MaartenGr/KeyBERT, *archived at* [https://web.archive.org/web/20250927082041/https://github.com/MaartenGr/KeyBERT]

Maarten 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

James W. Pennebaker, *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

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

Jonathan Shen, *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

UNUSED Jessica P. Ridgway, *et al.* 2021. Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study. *JMIR Medical Informatics* 9, 3, Article e23456 (Mar. 2021), 10 pages. https://doi.org/10.2196/23456 UNUSED

Dror Ben-Zeev, *et al.* 2013. Mobile technologies among people with serious mental illness: opportunities for future services. *Administration and Policy in Mental Health and Mental Health Services Research* 40, 4 (Jul. 2023), 340-343. https://doi.org/10.1007/s10488-012-0424-x

Chirag I. Patel, *et al.* 2012. Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study. *International Journal of Computer Applications* 55, 10 (Oct. 2012), 50-56. https://doi.org/10.5120/8794-2784

JaidedAI. 2024. EasyOCR: Ready-to-use OCR with 80+ supported languages and all popular writing scripts including: Latin, Chinese, Arabic, Devanagari, Cyrillic, etc. *GitHub*. Retrieved Nov. 25 2025 from https://github.com/JaidedAI/EasyOCR, *archived at* [https://web.archive.org/web/20251012081255/https://github.com/JaidedAI/EasyOCR]

Victor Sanh, *et al.* 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint.* https://doi.org/10.48550/arXiv.1910.01108

Liu Yinhan, *et al.* 2019. Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint.* https://doi.org/10.48550/arXiv.1907.11692

Jacob Devlin, *et al.* 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 1 (Jun. 2019), 4171-4186. https://doi.org/10.18653/v1/N19-1423

Mary Cece Young, *et al.* 2018. The Effects of Text-to-Speech on Reading Outcomes for Secondary Students With Learning Disabilities. *Journal of Special Education* Technology 24, 2 (Jul. 2018), 80-91. https://doi.org/10.1177/0162643418786047

Coqui.ai. 2025. Coqui.ai TTS. *TTS 0.22.0 documentation*. Retrieved Nov. 26 from https://docs.coqui.ai/en/latest, *archived at* [https://web.archive.org/web/20250812191341/https://docs.coqui.ai/en/latest]

pyttsx3.readthedocs.io. 2025. pyttsx3 - Text-to-speech x-platform. *pyttsx3 2.6 documentation*. Retrieved Nov. 26 2025 from https://pyttsx3.readthedocs.io/en/latest, *archived at* [https://web.archive.org/web/20251010005655/https://pyttsx3.readthedocs.io/en/latest]

Michael Röder, *et al.* 2015. Exploring the Space of Topic Coherence Measures. *In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)* (Feb. 2015), 399-408. https://doi.org/10.1145/2684822.2685324

Peter J. Kincaid, *et al.* 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Technical Report Research Branch Report, Millington, TN: Naval Technical Training*. Article RBR875, 51 pages. https://doi.org/10.21236/ADA006655

ITU-T P.800. 1996. Methods for subjective determination of transmission quality. *International Telecommunication Union*. 37 pages.

Réjean Plamondon and Sargur N. Srihari. 2000. On-line and off-line handwriting recognition: a comprehensive survey. IEEE Trans Pattern Anal Mach Intell (T-PAMI). *IEEE transactions on pattern analysis and machine intelligence* 22, 1, 63-84. https://doi.org/10.1109/34.824821

Cheng Cui, *et al.* 2025. PaddleOCR-VL: Boosting Multilingual Document Parsing via a 0.9B Ultra-Compact Vision-Language Model. *arXiv preprint.* https://doi.org/10.48550/arXiv.2510.14528

Li, M. et al. 2021. *TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models*. arXiv preprint. https://doi.org/10.48550/arXiv.2109.10282

Bai, J., et al. (2024). *Qwen-VL: A Frontier Large Vision-Language Model.* arXiv preprint. https://doi.org/10.48550/arXiv.2308.12966

Zuiderveld, K. (1994). *Contrast Limited Adaptive Histogram Equalization*. In Graphics Gems IV, Academic Press, 474–485. https://dl.acm.org/doi/10.5555/180895.180940

Otsu, N. (1979). *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076