

# My InsightOut

## DATA REPORT

### MonkeyBrains



# About us

**We at MonkeyBrains believe in the importance of implementing safe artificial intelligence (AI) in every organization that wants it. Having been educated at Radboud University, which has a strong history of promoting academic and industrial partnerships, we want to increase the connection between academic research, professional implementation, and corporate utilization.**

**Our vision is an ecosystem where automation and data-driven techniques can be applied in all fields, lessening workload, increasing comfort, and future-proofing systems. Our passion particularly extends to education, where we believe these advanced techniques can improve and revolutionize the field.**

**We are uniquely equipped with a blend of academic rigor and industry experience. With robust backgrounds in scientific research and practical applications, we can delve deep into complex subjects and implement cutting-edge techniques. This dual perspective allows us to bridge the gap between theory and practice, ensuring that our solutions are both innovative and practical.**





# Table of contents

<b>Introduction</b>	1
<b>Objectives</b>	2
<b>Data assessment</b>	3
<b>Core strengths of the dataset</b>	3
<b>Key challenges and limitations</b>	4
<b>Potential Analyses</b>	4
<b>Advantages</b>	5
<b>Disadvantages</b>	6
<b>Recommendations</b>	6
<b>Summary</b>	7
<b>Product Development</b>	8
<b>Conversational bot/ Robot Mirjam &amp; Magreet</b>	8
<b>AI-Driven Longitudinal Analysis: The AI Diary</b>	9
<b>Automated Data Cleaning and Theme Extraction</b>	9
<b>Enhancing Accessibility and Comfort</b>	10
<b>Conclusion</b>	11
<b>Bibliography</b>	12

# Introduction

*This report presents an in-depth analysis of my Insight Out's dataset, examining its potential for commercialization through direct sale to third parties or the development of proprietary products. Compiled by MonkeyBrains, the report explores the strategic value of the company's extensive knowledge base, assessing viable pathways for monetization and collaboration while providing well-founded recommendations for its optimal utilization.*

**my Insight Out (mIO)** is a digital platform dedicated to fostering self-reflection, self-knowledge, and personal growth. Built upon the scientifically validated Self-Knowledge Method (SKM) - previously known as the Self-Confrontation method (SCM) - this innovative approach enables individuals to examine the deeper meaning behind their life experiences, empowering them to navigate personal development with greater awareness and intentionality. At its core, my Insight Out is inspired by the Valuation Theory of renowned psychologist Prof. Hubert Hermans, who posits that individuals possess unique self-knowledge and the ability to shape their own growth (Hermans, 1991). By leveraging this theory, mIO provides a structured yet deeply personal framework for self-inquiry, helping users uncover unconscious patterns, gain valuable insights, and ultimately take ownership of their own development.

What makes my Insight Out truly unique is its ability to function as both a personal tool and a collective knowledge system. With roots in over four decades of coaching and research, the platform draws upon an extensive database of self-reflective experiences, forming a vast repository of human insight. This data—accumulated from thousands of users engaging in structured self-exploration—offers profound potential, both in terms of its individual impact and its broader applicability. The platform's approach aligns with emerging scientific and philosophical perspectives that posit the self as a dynamic, evolving process rather than a fixed entity. Through its integration of the Extended Self concept from Valuation Theory, mIO positions personal experiences within a collective framework, suggesting that individual growth is enriched by the shared narratives and reflections of others.

Beyond its role as a transformative self-development tool, the curated interview and corresponding user data may hold potential for third parties. The structured nature of the platform, combined with its deep psychological foundations, has led to the creation of a unique and highly organized dataset—one that could hold commercial and academic interest. The question now arises: should this data be further developed into a marketable product, or does its greatest potential lie in direct partnerships with third-party organizations? This report will explore the possibilities, assessing the feasibility of commercializing mIO's dataset, identifying potential applications, and outlining strategic recommendations for its future.

## Objectives

The objectives of this report, which have been determined contractually, are elaborated upon below, and have been expanded to a few additional general insights, based on the statement that “the Contractor must not withhold any findings and research from The Client”.

### **Data assessment:**

The critical component of this study is a thorough evaluation of mIO's data to determine its composition, quality, and potential value. This assessment will include an analysis of how the data has been obtained, organized and stored, with a special focus on what components of the data are yet to be uncovered, which might hold crucial value. Specifically, we outline the core strengths of the data, its challenges and limitations, and improvement considerations.

### **Product development:**

Beyond monetizing data through direct sales, the study will explore opportunities for mIO to develop innovative products leveraging their dataset. Potential applications include AI-driven self-reflection tools, advanced coaching platforms, and potential statistical analyses to provide deeper psychological insights. The report will identify promising product concepts, evaluate their feasibility, and recommend implementation strategies.

The report will follow the above structure, using the data assessment to ground the evaluation of possible data sales and product developments. We conclude with a summary of the key points discussed and an overall assessment of the marketability of the proposed solutions.

## Data assessment

The first step in determining an appropriate path, between selling the data or product development, is analyzing the data that we have available. After careful consideration of the provided documents - consisting of an overview of the database structure and several API responses - we provide an analysis of what we consider to be the core strengths of the dataset, as well as its main limitations. We propose actionable recommendations for better understanding of the dataset, as well as improvements on future data collection.

The data primarily consists of 24 emotions ranked in intensity, core sentences that elicited these emotions, perceived themes between core sentences and correlated responses, and strategies that address these themes. The presence of these being dependent on which dataset the point lies in, either DB1 or DB2. This multimodal data allows possible underlying relationships to be established between the numerical and natural language formats available. To further uncover what value lies in the data, further analyses can be carried out to understand these hidden relationships between the stories/themes, strategies, and the emotions that underlie them. These analyses can be either carried out by mIO with the help of third-parties or the potential analyses can be marketed to customers as the source of value itself. This will be explored further in the potential analyses section.

### Core strengths of the dataset

The dataset's most compelling advantage lies in its **specialized framework for emotional quantification**. This pairs a formal model for mapping emotions to sentences, with a methodology of introspection through reflecting on a core experience. Assigning more structured emotional chords to these subjective core experiences, in addition to capturing the contextual layers of individual bias, coping mechanisms, and personal themes.

Additionally, the dataset's curation and review process ensures that it is relatively clean, with 3,000 manual entries (and 6,000 across combined databases), offering consistency, which is rare in qualitative emotional and natural language data in general. A unique differentiator is the linkage of emotions to **core human experiences**—specific life events or themes (e.g. loss, achievement) and their associated emotional responses. This multi-layered approach, absent in existing datasets, offers potential across several industries, from marketing firms to consumer lifestyle applications.

**mIO vs existing sentiment analysis** From a content perspective, the data contains a unique offering. As mentioned in the first paragraph, we have a formal model for mapping sentences to implicit emotions. Existing datasets that have this capability fall under “sentiment analysis” (Wankhade et al., 2022). However, these datasets, which are human-curated, often only contain ratings on a unidimensional scale between highly positive and highly negative or even simple binary positive/negative choices (Tanana et al., 2021). Therefore, these datasets only have a single output available whereas the mIO dataset contains responses with 24 specific emotions rated from 1-6. This substantially increases the emotional understanding of the core experiences compared to the standard responses available in the field.

## Key challenges and limitations

Despite its strengths, the dataset faces significant hurdles. Its **limited sample size** (~6,000 entries) poses risks for machine learning applications, as modern AI models often require vast training data to generalize effectively. Furthermore, the intentional ambiguity of responses, while philosophically aligned with the project’s goals, complicates integration with natural language processing (NLP) systems. For instance, sentiment analysis algorithms struggle with abstract or context-dependent phrasing, which could hinder automated analysis.

Another critical gap is the **lack of demographic metadata** such as age, cultural background, and gender. While omitting this data reduces privacy risks, it also limits the ability to analyze how emotional responses vary across populations—a key consideration for fields like marketing or mental health. Without such context, the dataset risks oversimplifying human emotion as a universal construct, overlooking cultural or age-specific nuances and third-party interest.

## Potential Analyses

Below is a list of analyses which can be conducted to better understand the data, either by mIO or by the prospective customers for the dataset.

### Means and Variances

Basic analysis of the means and variances of emotions across responses can also lead to understanding what emotion levels typify a core experience.

### **Correlations**

Pairwise correlations between emotions across responses can be carried out to see if some emotions are redundant or anti-correlated, leading to a better understanding of what a natural emotion set is.

### **Clustering**

The emotional scores can be clustered along a 24 dimensional "emotion space" which can lead to discovering a simpler categorization of emotional stereotypes that appear more often than others, allowing experiences to be classified into these stereotypes

### **Dimensionality reduction**

Similar to the pairwise correlations, it could be that there is a lower dimensional "feeling space" which underlies the 24 dimensional emotion space, leading to further understanding of the dimensionality of emotions (van der Maaten, Postma & van den Herik, 2009)

### **NLP Analyses**

Standard NLP techniques such as "bag of words" analyses can be carried out to understand which words and phrases in core sentences and themes correspond to which emotional sets. (Atkins, D.C. and Imel, Z.E., 2021)

### **Sentiment Analysis**

Pretrained language models can be further trained by this data to understand the relationship between emotions and speech but this may be limited by the small sample size.

There are a number of technical advantages and disadvantages to the data in both its format and content. Below we list some of these considerations that a potential customer would find within the dataset.

## **Advantages**

### **High Dimensional**

The emotional data is high dimensional and therefore provides rich information within each response.

### **Multimodal**

there are both numerical and natural-language responses for each experience in the dataset and that provides a natural labeling for machine learning tasks.

## Disadvantages

### **Small sample size**

The entire dataset only consists of 6000 cleaned samples in total, where only a subset of these are complete. With some entries missing interpretations of correlated responses, identified themes, or strategies.

### **Lack of specificity**

The core experiences are intentionally post-processed to be less personal and more general leading to less emotional valence left to be uncovered by NLP algorithms.

### **Lack of demographics**

Besides specific gender information in a subset of the dataset, the use of the dataset becomes constrained as it does not have the demographic data that would allow customers to use it for specific targeting.

While this report will not go into the specifics of how this technology works, a short section on the technical aspects is felt to be insightful.

## Recommendations

Based on the assessment of the data, we make one recommendation for improving the overall viability of the dataset and propose two potential selling points that can be highlighted while devising a strategy:

### **Metadata Enrichment**

Collect anonymized demographic data using privacy-preserving techniques (e.g., aggregated age ranges rather than exact ages). This would enable segmentation for targeted applications without compromising user anonymity.

### **Professionally Curated Dataset**

The relative cleanliness and standardization of this data, having been generated under the guidance of a clinician, positions it as both unique and valuable. We believe that this is a selling point that can be emphasized when marketing the data. Highlighting the manual curation work done by experts such as Mirjam and Magreet differentiates the dataset from unverified and less consistent alternatives.

### **Multimodal Nature**

Furthermore, the quantitative aspects of the dataset should be emphasized as it disambiguates the typically vague emotions around a natural language response. This also sets apart the mIO data from the typical dataset which will often contain either only the quantitative metrics or the natural language responses.

### **Summary**

Overall, the dataset offers a unique, multimodal, high-dimensional, set of responses - yet is lacking in size and specificity. The unique methodology used to generate the data gives a highly specialized set of responses compared to the norm. Additionally, the high dimensionality and multimodal nature can allow for many possible analyses to be carried out. The use in large AI models is limited, however, by the small sample size and ambiguity of the core sentences. The drawbacks may be limited by collecting further demographic data for each user along with the specificity of the core experiences being left intact. It is understood that these changes may lead to privacy and anonymity concerns for the users, so these changes are recommended with caution.

## Product Development

Another potential avenue for leveraging this dataset or process is to create a stand-alone product. This product can be sold directly to end-users or accessed by third-party clients, as discussed in the previous section. A significant strength of the platform is the clarity provided by the conversational partners, Mirjam and Magreet. Their input enriched the process of completing the mIO questionnaire by guiding users to focus on the psychologically relevant aspects of their answers.

The unique selling point of the process is its human-curated design. With advancements in large language models (LLMs), it is now feasible to automate the role of the conversational partner. Alternatively, a variant of the mIO process could be offered through an AI-Diary application. We explore these product ideas based on mIO and additional ways to streamline the process itself.

### Conversational bot/ Robot Mirjam & Magreet

One way of extending the current mIO process lies in the development of a conversational bot, a tool designed to transform static questionnaires into dynamic, empathetic dialogues. Unlike traditional surveys that ask users to rate 24 emotions against a rigid prompt (“How much of this emotion does this situation make you feel?”), the bot would replicate the nuanced guidance provided by human facilitators like Mirjam and Magreet. By engaging users in natural conversation, it would extract emotional values organically, preserving the “closeness” of human interaction while ensuring data homogeneity.

The bot’s strength would stem from its ability to **adaptively clarify questions** and **steer users toward productive insights**. For example, if a user provides a superficial response, the bot could ask follow-up questions like, “What underlying qualities of this situation made you feel that way?”—mirroring the introspective rigor of human curators. Fine-tuning the model to adhere strictly to the system’s methodology would prevent conversational drift, maintaining both contextual coherence and data cleanliness.

#### **Leverage NLP frameworks**

(e.g., GPT-4, Deepseek) to train the bot on Mirjam/Magreet’s methodology, ensuring adherence to structured emotional prompts.

#### **Design conversational guardrails**

(e.g., scripted follow-ups like, “Can you describe the source of that emotion?”) to prevent off-topic digressions, or leveraging internal prompt summary to scan user and model output for relevance (e.g., Lang-chain, Pydantic-ai)

### **Beta-test with existing users**

to refine dialogue flow, prioritizing responses that mirror human facilitator success rates.

## **AI-Driven Longitudinal Analysis: The AI Diary**

A **weekly/monthly AI Diary** feature could complement the conversational bot by identifying patterns in emotional responses over time. For instance, it might flag recurring stressors tied to specific emotional “chords” (e.g., “Every Monday, frustration peaks due to workload”) and it can offer reflective prompts like, “How might shifting your coping strategy alter this pattern?” This tool could serve as a diagnostic aid for therapists, providing structured insights into a client’s emotional trajectory without replacing human judgment.

### **Build pattern-detection algorithms**

(e.g., GPT-4, Deepseek) to train the bot on Mirjam/Magreet’s methodology, ensuring adherence to structured emotional prompts.

### **Create therapist dashboards**

(e.g., scripted follow-ups like, “Can you describe the source of that emotion?”) to prevent off-topic digressions, or leveraging internal prompt summary to scan user and model output for relevance (e.g., Lang-chain, Pydantic-ai)

### **Partner with mental health apps**

to test the diary as a supplemental diagnostic tool.

## **Automated Data Cleaning and Theme Extraction**

An **automated sentence cleaning process** would streamline the curation of user responses. By simplifying entries to their core themes (e.g., distilling a 200-word reflection into “fear of failure linked to childhood expectations”), the system would ensure maximal analytical utility while preserving emotional nuance. This process could mimic the manual cleaning done by Mirjam and Magreet, applying rules like:

- **Removing redundant or off-topic tangents.**
- **Tagging responses with relevant emotional chords.**
- **Flagging ambiguous entries for human review.**
- **Use rule-based sectioning**  
(e.g., keyword matching for “feelings” or “situations”) to auto-label entries with emotional chords.

- **Train ML models**

to detect and remove redundancies (e.g., repeated phrases) and implement abstraction while preserving core themes.

- **Deploy hybrid review systems**

AI handles 80% of cleaning, while ambiguous cases are routed to human curators.

## Enhancing Accessibility and Comfort

To deepen user engagement, auxiliary features like **voice-to-text/text-to-voice** and customizable avatars could be integrated. While non-essential, these elements would humanize the experience, making users feel heard rather than interrogated. A warm, conversational tone and paced dialogue would further reduce friction, encouraging honesty and reducing survey fatigue—a common issue in emotion-tracking tools.

### **Integrate voice-to-text APIs**

(e.g., Google Speech-to-Text) with emotion-aware pacing (e.g., pauses after sensitive responses).

### **Develop minimalist avatars**

with subtle, non-distracting animations (e.g., nodding) to signal active listening.

### **Implement a “comfort score” feedback loop**

where users rate interactions post-session to iteratively refine tone and pacing.

## Conclusion

mIO presents a unique and valuable resource for understanding human emotion, derived from a robust, psychologically grounded methodology. While the dataset itself holds significant potential for commercialization, its limitations in size and specificity pose challenges for immediate large-scale AI adoption. Therefore, a dual approach is recommended.

First, mIO should prioritize enriching the existing dataset through metadata enhancement, collecting anonymized demographic information, and potentially relaxing some constraints on core experience specificity, while carefully considering privacy implications. This will increase its appeal to potential buyers, who can leverage the data for deeper consumer insights, targeted marketing, and understanding emotional responses to specific life events. The professionally curated nature of the dataset, combined with its unique methodology, should be strongly emphasized in marketing efforts.

Second, mIO should focus on developing its proprietary technology into marketable products. Building upon the expertise of its creators, the development of an AI-driven Conversation Bot offers a compelling way to scale the platform's unique interview process while maintaining data quality. This, combined with features like an AI Diary for longitudinal analysis and automated data cleaning, creates a suite of tools with strong potential in the burgeoning mental health and lifestyle application market. These products can be offered directly to consumers or through partnerships with existing platforms. This strategy allows mIO to capitalize on its core strengths – its unique methodology and the expertise of its creators – while mitigating the limitations of the current dataset.

From our perspective, the dataset holds significant potential but has limited value in its current form. This limitation can be addressed by either increasing the diversity of the data or by developing a product based on the dataset. Practically, we believe that creating a product is the more feasible option, as enhancing the dataset would require the creation of a new dataset with additional metadata.

## Bibliography

Hermans, H. J. M. (1991). The person as co-investigator in self-research: Valuation theory.

*European Journal of Personality*, 5(3), 217–234.

<https://journals.sagepub.com/doi/10.1002/per.2410050304>

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges.

*Artificial Intelligence Review*, 55(7), 5731–5780.

<https://link.springer.com/article/10.1007/s10462-022-10144-1>

Sukhavasi, S., Sharma, K. and Kodati, P., 2023. Emotional Marketing on Consumer Behaviour - Perception Study.

*ResearchGate*

<https://www.researchgate.net/>

[publication/369451859\\_EMOTIONAL\\_MARKETING\\_ON\\_CONSUMER\\_BEHAVIOUR\\_-\\_PERCEPTION\\_STUDY](https://www.researchgate.net/publication/369451859_EMOTIONAL_MARKETING_ON_CONSUMER_BEHAVIOUR_-_PERCEPTION_STUDY)

Tanana, M.J., Soma, C.S., Kuo, P.B., Bertagnolli, N.M., Dembe, A., Pace, B.T., Srikumar, V., Atkins, D.C. and Imel, Z.E., 2021. 'How do you feel? Using natural language processing to automatically rate emotion in psychotherapy'.

*Behavior Research Methods*, 53(5), pp.2069-2082

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8455714/>

van der Maaten, L., Postma, E. and van den Herik, J., 2009. 'Dimensionality Reduction: A Comparative Review'.

*TiCC TR 2009–005, Tilburg University*

[https://members.loria.fr/moberger/Enseignement/AVR/Exposes/TR\\_Dimensiereductie.pdf](https://members.loria.fr/moberger/Enseignement/AVR/Exposes/TR_Dimensiereductie.pdf)

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D. and Pappas, I., 2023. A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection'.

*Artificial Intelligence Review*, 56(1), pp.1-50.

<https://link.springer.com/article/10.1007/s10462-023-10509-0>