Check for updates

RESEARCH ARTICLE

The Limits of AI in Understanding Emotions: Challenges in Bridging Human Experience and Machine Perception

[version 1; peer review: awaiting peer review]

Alain Finet, Kevin Kristoforidis, Julie Laznicka

Health Institute, Financial Management Department, University of Mons, Mons, Walloon Region, Belgium

v1 **First published:** 13 Jun 2025, **14**:582
https://doi.org/10.12688/f1000research.164796.1

Latest published: 13 Jun 2025, **14**:582
https://doi.org/10.12688/f1000research.164796.1

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background

This article explores the potential of Artificial Intelligence (AI) to replace the human factor in the analysis of emotions within financial decision-making contexts. The background of this study is as follows: growing reliance on Artificial Intelligence for text analysis, understanding its capacity to detect emotional patterns is particularly relevant in fields where emotional dynamics significantly influence behavior, such as trading.

Method

Our research method is based on a three-day trading experiment involving students, during which participants made decisions under conditions of uncertainty. At the end of the experiment, each participant wrote a free-form report describing their emotional experience. We applied four different Artificial Intelligence tools to analyze these texts and compared the results with those obtained through a traditional lexical analysis method. AI tools provided relatively consistent and coherent assessments of the general emotional tone in the texts.

Results

Our results suggest that they successfully identified dominant emotions such as fear, disappointment, and hope. However, the ability of Artificial Intelligence to detect more nuanced or mixed emotional states was limited. In contrast, traditional lexical analysis offered greater sensitivity to emotional complexity. Although both approaches converged on broad emotional tendencies, divergences appeared when addressing subtle or context-specific emotional shifts. Our findings suggest that AI can serve as a useful preliminary tool in the analysis of emotions in written texts, especially for identifying dominant patterns.

Conclusion

In conclusion, according to us, due to its limitations in detecting emotional nuances, Artificial Intelligence should be integrated into a hybrid analytical approach. Combining Artificial Intelligence with traditional or human-led analysis enhances both the precision and depth of interpretation. We argue that a hybrid model avoids the oversimplification of emotional data and better reflects the complexity of emotional dynamics in decision-making under uncertainty.

Keywords
emotion analysis, qualitative research, stock markets, Artificial Intelligence, lexical analysis.

Corresponding author: Julie Laznicka (julie.laznicka@umons.ac.be)

Author roles: Finet A: Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Kristoforidis K**: Data Curation, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Laznicka J**: Data Curation, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research received financial support from the Walloon-Brussels Federation (Belgium) through a Concerted Research Action (Grant Number: ARC-25/29 UMONS5).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Finet A, Kristoforidis K and Laznicka J. **The Limits of AI in Understanding Emotions: Challenges in Bridging Human Experience and Machine Perception [version 1; peer review: awaiting peer review]** F1000Research 2025, **14**:582 https://doi.org/10.12688/f1000research.164796.1

First published: 13 Jun 2025, **14**:582 https://doi.org/10.12688/f1000research.164796.1

Introduction and state of the art

Emotion recognition is a widely analyzed issue in recent literature. From the survey of 142 journal articles by Khare et al. (2024),[1] it appears that models of emotion understanding generally take into account physical (facial and vocal recognition) and physiological (electroencephalogram, electrocardiogram, skin sudation, heart rate and eye tracking) signals, in different contexts: education and pedagogical strategies,[2,3] the medical field,[4] in the industrial sector,[5,6] etc. Several authors emphasize the need to use different measurement instruments simultaneously in order to approach the psychological and emotional reality of individuals.[7]

Unlike the analysis of vocal speech and micro-expressions on the face, the topic of textual analysis and the emotions it carries is rarely addressed in scientific studies.[8] The studies (see,[9] for an exhaustive list of techniques) focus mainly on the classification of emotions or the level of polarity of emotions with different scales of measurement (e.g. very positive, positive, neutral, negative, very negative). Deep learning and machine learning seem to be the most commonly used methods for this issue.[10] According to the results of the study by Machová et al. (2023), the combined use of 1D convolutions (to extract local characteristics) and LSTMs (Long Short-Term Memory) achieves 91% accuracy in the classification of six basic emotions.[11] In addition, Bharti et al. (2022) demonstrate that the combined use of deep learning (e.g., convolutional neural networks) and machine learning techniques resulted in high levels of accuracy in the assessment of emotions (happiness, disgust, fear, surprise, anger, and sadness).[12] In order to assess emotions and their intensity (positive, negative and neutral) in written documents, the effectiveness of the combined use of deep learning and machine learning has also been confirmed.[13] Furthermore, in their survey of the different measurement instruments, Murthy and Kumar (2021) emphasize the importance of comparing and combining computational and lexical approaches for the detection of emotions from textual data.[14]

Approaches focusing on Artificial Intelligence in the field of qualitative research seem less present in literature, which can largely be explained by the recent nature of their use. As qualitative research is based on the reduction of data into codes and categories that facilitate synthesis and interpretation, Artificial Intelligence could be a tool to help generate codes, given the appropriate control of prompts and parameters to adjust requests to the orientation of the research.[15] In addition, analyses emphasize the importance of prompt designs (for example, in the case of Amazon Mechanical Turk, a question asking whether a word is associated with an emotion leads to better results than asking whether a word evokes an emotion),[16] the accurate presentation of the analysis context, the focus on the chosen methodological perspective as well as the analytical process and the definition of data formats.[17]

Contributions from qualitative research[18,19] highlight time savings through the use of Artificial Intelligence (for example, in comparison with the time required for manual coding and following a rigorous analysis procedure)[20]; on the other hand, artificial intelligence would be more effective for the identification of specific themes that are not subject to interpretation than more subtle themes. Artificial intelligence would therefore be mainly useful in the early stages of analysis (for example, the creation of causal link diagrams).[21] In the recognition of emotions, given their complexity and their nuanced nature, the unclear nature of emotional boundaries and the incompleteness of emotional information, this latter observation is of main importance, which complicates the task of assigning a specific emotional category.[22]

One of the limitations of Artificial Intelligence is its inability to generate emotion,[23] so the question that naturally comes to mind is its ability to perceive, name and even measure emotions. Thus, even if the use of Artificial Intelligence is an option for qualitative research, different authors question its place in the general methodological approach and insist on the role of the human factor in monitoring and interpretating the results.[24] There is no optimal method applicable to all contexts and, in all cases, careful reading and rigorous validation by researchers would also be necessary.[25] Many authors have emphasized the need for a hybrid approach, combining Artificial Intelligence and analysis using human methods.[26–29]

Based on these different strengths (saving time, prior analysis of large amounts of data) and weaknesses (the difficulty of addressing the nuance of subtle elements), our research questions are as follows:

- Do Artificial Intelligences make it possible to identify the emotional patterns in written documents in a uniform manner, regardless of the Artificial Intelligence selected?
- Do Artificial Intelligences reach the same conclusions as traditional lexical measurement tools?

Sampling, materials and methods

Sampling and general methodological orientation

In order to answer the two research questions above, we will take into consideration written data by eight students studying Management Sciences at the University of Mons (Belgium) who took part in an experiment in the field of trading over three consecutive days (from January 27 to 29, 2025). All participants were adults (aged 18 years or older) and provided written informed consent prior to participating in the study. Consent was obtained using printed forms (the consent form is available in the "Data Availability" section), in accordance with ethical guidelines for research involving human participants. Although the experiment was conducted on a student population (all students were over 18 years of age), no manipulation was carried out: in practical terms, they remained seated in front of a computer for several hours without any physical interaction with the organizers. Finally, no intrusive technology was used, and no neurophysiological measurement tools were used. For all these reasons, how the experiment was designed does not fall within the scope of the Helsinki guidelines. Helsinki guidelines concern medical research involving human participants, but in our case, this is not medical research at all, but simply using written documents.

Using student populations in the field of experimental finance seems to make sense.[30–33] Referring to the study by Dorn and Sengmueller (2009), we keep in mind, however, that students not directly interested in the financial value of portfolios may tend to overplay.[34] The adaptations presented below, however, are made to moderate this limitation.

Financial constraints mainly determined the number of students selected (students were paid to take part in the experiment). Over the three days, the students were asked to place orders on the basis of a portfolio consisting of 40 French stocks (CAC40 index) and a virtual portfolio of 100,000 euros. No instructions were given by the organizers on the number of transactions to be carried out, nor in terms of their volume. It was also planned to reward the highest financial portfolio in order to highlight the development of differentiated emotional patterns over time. The various experiments conducted since 2019 on the issue of decision-making on the stock markets have given us the opportunity to consolidate our experimental protocols.[35,36]

In terms of methodology, we use an analysis from experimental finance and supported by qualitative analysis tools. This double orientation reflects an evolution in the field of financial research. Indeed, in the '80s, Grether (1981) noted that experimentation was generally conducted by researchers in psychology but very little in finance.[37] Under the influence of behavioral finance and better control of some technological tools, research methods have tended to shift.[38] Bloomfield and Anderson (2009) also argue that experimentation is underused in the financial field but would be useful for testing behavioral finance and biases.[39] Controlled manipulations would also have the advantage of building an environment in which a causal theory of phenomena could be assessed with a high level of validity.[40]

The material collected at the end of this experiment included trading notebooks in which the different transactions carried out and their financial volume were recorded, data collected through participant observation, semi-structured interviews, as well as written documents at the end of the experiment aimed at presenting how the experiment had been *emotionally* experienced by the participants. For answering our research questions, we will mainly focus on this last material.

Stock market context and trading strategies

The experiment took place during a bear market, even if the losses were relatively small, as shown in Table 1.

Through participant observation, we noticed that the students took a very strong personal interest in the (virtual) portfolio. In other words, they seemed to be emotionally affected by the negative variations in the financial value and, as they experienced disappointment after disappointment, a feeling of abandonment seems to have developed (some students hardly placed orders at all on the last day). This finding is also supported by the trend in the number of transactions over the three days (Table 2).

Through participant observation, we noticed that the students took a very strong personal interest in the (virtual) portfolio. In other words, they seemed to be emotionally affected by the negative variations in the financial value and, as they experienced disappointment after disappointment, a feeling of abandonment seems to have developed (some students

Table 1. CAC 40 Index, DJ30 Index, NASDAQ 100 Index and TOPIX Index performance over the experimentation period.

Date	January 27, 2025	January 28, 2025	January 29, 2025	Total
CAC 40	-0,0003	-0,00012	-0,0032	-0,0036
DJ 30	0,0065	0,0031	-0,0031	0,0065
NASDAQ 100	-0,0297	0,0159	-0,0024	-0,0162
TOPIX	0,0026	-0,0004	0,0068	0,009

Table 2. Trading strategy for three days of experimentation.

	Total D1[1]	B[2]	S[3]	Total D2	B	S	Total D3	B	S	Total	B	S
I.1.	21	12	9	18	15	3	8	5	3	47	32	15
I.2.	28	17	11	8	4	4	7	3	4	43	24	19
I.3.	17	15	2	4	1	3	2	1	1	23	17	6
I.4.	49	28	21	14	10	4	12	9	3	75	47	28
I.5.	17	11	6	20	10	10	16	7	9	53	28	25
I.6.	13	10	3	2	1	1	0	0	0	15	11	4
I.7.	13	9	4	9	6	3	1	0	1	23	15	8
I.8.	11	10	1	9	3	6	16	8	8	36	21	15
Total	169	112	57	84	50	34	62	33	29	315	195	120
Mean	21,12	14	7,12	10,5	6,25	4,25	7,75	4,125	3,62	39,37	24,38	15

[1]Day.
[2]Buy.
[3]Sell.

Table 3. Participants' ranking at the end of the experience.

Ranking	Participant	Portfolio value	Change in %	Net change in %[1]
1	I.7.	100213,73	0,21%	0,57%
2	I.3.	99887,01	-0,11%	0,25%
3	I.6.	99394,53	-0,61%	-0,25%
4	I.4.	99314,31	-0,69%	-0,33%
5	I.2.	98400,4	-1,60%	-1,24%
6	I.1.	98249,79	-1,75%	-1,39%
7	I.8.	97850,6	-2,15%	-1,79%
8	I.5.	97503,76	-2,50%	-2,14%

[1]Portfolio Change – CAC40 Change.

hardly placed orders at all on the last day). This finding is also supported by the trend in the number of transactions over the three days (Table 2).

It should be noted that at the end of the three trading days, seven out of eight portfolios lost financial value, but two of them outperformed the CAC 40 index (Table 3).

In a bear market, the question is no longer about being the best, but about being less bad than the others. In other words, the reward would be achieved through a passive strategy, since the risks taken seem to result in financial losses. This behavior also helps to limit regrets. The accumulation of negative experiences does indeed cause emotional discomfort and withdrawal.[41] Based on the combined findings of the participant observation and the development of a withdrawal strategy, we postulate that the emotions developed must have had negative charges.

Selected analysis tools

As specified in the methodological presentation, at the end of the three-day experiment, the participants were asked to write down the moments perceived as decisive and central for them, as well as all the events that had affected them, both positively and negatively. The participants were free to choose the format, length and degree of detail of their explanation (see Table 4). Our approach is in line with different authors who demonstrate the importance of written documents for understanding emotions and their effects.[42,43] The emotional patterns expressed in the comments will be identified by four artificial intelligences (ChatGPT, Grok, DeepSeek and Yiaho) and by a lexical approach involving only human intervention. The use of a lexicon to identify emotions from textual data is one of the main analysis techniques for emotion recognition.[44]

Table 4. Descriptive statistics of written documents analyzed.

	Number of Documents	WD 1, Words	WD 2, Words	WD 3, Words	WD 4, Words	WD 5, Words	WD 6, Words	WD 7, Words	Mean
I.1.	6	42	36	25	50	37	25		35,833
I.2.	7	65	76	40	65	40	39	13	48,286
I.3.	4	33	14	7	16				17,500
I.4.	3	31	26	17					24,667
I.5.	7	63	60	35	39	35	38	17	41,000
I.6.	5	26	34	29	21	41			30,200
I.7.	6	59	65	32	30	19	24		38,167
I.8.	7	36	34	42	40	35	31	31	35,571
Total	45								

Lexical approaches are based on an analysis of the words from the text, each of which is assigned a score indicating whether it can possibly be linked to a given emotion (0 means no association and 1 implies an association with one or more emotions). Among existing lexicons, the NRC Emotion Lexicon and DepecheMood++ are the most frequently used for emotion detection from text data.[45] We selected the NRC Emotion Lexicon because it covers a wider range of emotions than DepecheMood++, which is limited to the six basic emotions defined by Ekman (1992).[46] The NRC Emotion Lexicon is based on Plutchik's theory (1980) and classifies words according to eight emotions: anger, fear, anticipation, disgust, happiness, sadness, surprise and confidence.[47] It also includes a valence dimension (positive or negative) that indicates an emotion's level of pleasure in accordance with Russel's model,[48] thus offering another perspective for analyzing the emotions detected in a text.[49] From our perspective, valence can be associated with polarity.

Results

Descriptive statistics of written documents

Prompt selection procedure

Several prompts were developed in the analysis carried out by artificial intelligence to ensure the correct understanding of what was required. Thus, initially, we worked on the basis of the following prompt:

With reference to the following emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust), give me the emotional combination and the proportion related to these emotions considering the following sentences.

Artificial Intelligence (ChatGPT) provides emotional scores (percentages), explains the reasons underlying this choice in three or four sentences, produces a summary table of the results and a general conclusion.

The explanations, which are presented in detail for each of the emotions considered, seem logical and some nuances are also provided, for example:

> *This sentence is mainly focused on sadness and disappointment, with a slight touch of disgust and anger about the situation* (ChatGPT).

After reading the explanations provided by the Artificial Intelligence (ChatGPT) and related to the analysis of some written sentences, we noticed that, in the *mind* of Artificial Intelligence, the emotion linked to confidence could have both a positive connotation (an improvement in self-confidence) and a negative one (a loss of self-confidence), whereas, in the selected theoretical classification, confidence is only considered in the positive sense.

> *Confidence: the participant expresses a low level of confidence, particularly in the expertise of others, estimated proportion: 30%* (ChatGPT).

To overcome this limitation, we decided to replace confidence with optimism in the prompt, which was as follows:

Prompt 1: *Referring to the following emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, optimism), give me the emotional combination and the proportion related to these emotions considering the following sentence.*

Firstly, the results from this prompt resulted in total emotional percentages, in some cases above one and in other cases below one. Following these results, which seemed surprising to us, ChatGPT was asked to understand the reasons behind them. The explanations given by ChatGPT were as follows:

More than 100% (ChatGPT): *The sum of the percentages may exceed 100% in some cases due to the multiplicity of emotions felt simultaneously or empathy shared with different emotions. In the passages provided, it is possible that several emotions were present at different degrees at the same time, which explains why the percentages do not always exactly total 100%.*

Here are some reasons why this might happen:

1. Multiple simultaneous emotions: A complex emotional situation can include several simultaneous emotions. For example, a person may feel anticipation while also being surprised by an event that occurs. Each emotion is then analyzed separately, resulting in percentages that, when added together, exceed 100%.

2. Contradictory emotions: Sometimes a person can feel both positive and negative emotions, which can lead to a greater diversity in proportions. For example, fear or uncertainty can coexist with optimism or happiness, thus increasing the diversity of emotions and, consequently, of proportions.

3. Approximations and variations in the analysis: When analyzing emotions in a text, there may be subtle nuances in the way an emotion is perceived and expressed. This can lead to slight variations in the allocation of percentages to each emotion, which, in some cases, can result in the sum of the proportions exceeding 100%.

In the case of a result less than one (ChatGPT): *The sum of the percentages may be less than 100% in some cases for several reasons:*

1. Unexpressed emotions: If certain emotions are not present or have not been identified in a given passage, this can result in a sum of less than 100%. For example, if the passage does not mention any form of sadness or happiness, these emotions would not be considered in the analysis, which could lead to percentages of less than 100%.

2. Dominant emotions prioritized: During the analysis, certain emotions may dominate or be much stronger than others, so that only the main emotions are identified, and the weaker ones are left out. This can result in a sum that covers only part of the emotions present, with proportions not totaling 100%.

3. Emotional nuance: Some emotions may be subtle or not explicitly expressed in the text. For example, slight anticipation or optimism may not be strong enough to be fully counted, resulting in a sum of less than 100%.

4. Complexity of human emotion: Human emotions are often nuanced and can overlap in subtle ways. In some cases, some emotions may be too subtle or difficult to identify precisely to be counted, thus reducing the sum of the percentages.

5. Adjusting percentages: Sometimes percentages are adjusted to exact numbers, which can result in a total of less than 100% due to adjusting inaccuracies.

6. Distribution of emotions: If a passage expresses an emotion in a more general or indirect manner, it may not be attributed as much importance as other more obvious emotions, which can also contribute to a sum that is less than 100%.

To address this issue, ChatGPT offered the following answer: *To correct this phenomenon in a more rigorous analysis, it would be necessary to adjust the proportions so that their sum is 100%* (ChatGPT).

Secondly, ChatGPT placed emotions that it was unable to quantify in an “Other” category. In some cases, this represented more than 20% of the results and the percentages for this category exceeded the percentages for the other emotional labels. It should be noted that the other three Artificial Intelligences selected did not highlight this category.

Based on these two elements, we reformulated the prompt as follows:

Prompt 2: *Referring to the following emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, optimism), give me the emotional combination and the proportion related to these emotions. Considering the following sentence, give me the percentage of emotions not specified in the list and reach a total of 100%.*

Both prompts were used for each written document from each participant. Based on the results from the texts, emotions were identified, and we calculated averages for each emotion given by the Artificial Intelligences.

Results analysis

Results from Prompt 1

It should be noted that due to the many limitations related to prompt 1, we do not include here the lexical analysis in our results. It will only be carried out for prompt 2. To avoid making the paper too long, we decided to work in terms of emotional couples (see Table 5) which include the two most important emotional components for each Artificial Intelligence (in some cases, the percentages were similar and therefore three emotional labels were taken into consideration). In the appendix to the article, there is an example for one participant of the results from the first prompt (see Table 14).

Table 5 shows a lack of homogeneity in the results given by the Artificial Intelligences. Sadness seems to be much less present for ChatGPT than for Grok, DeepSeek and Yiaho. As mentioned, the category "other emotions" (main emotional component for I.3. and I.6., secondary for I.2.) could hide some emotional nuances not included in the ChatGPT results. This table also highlights the relative homogeneity between DeepSeek and Yiaho's results.

Considering all the emotional charges in the couples, sadness seems to be the emotion most commonly felt, which could be explained by the degraded context in which the experiment was organized (see Table 6). Some emotional combinations may seem surprising, for example (sadness, anticipation). The explanation we put forward for this example is that the expectations of some participants (the second component of the combination) were not met and this lack of satisfaction led to the development of sadness. One of the major limitations of Artificial Intelligence when considering a couple's approach is the difficulty in perceiving the extent to which one component could influence the other.

Negative emotions dominate our findings (see Table 7), even though the results related to positive emotional charges suggest that emotional complexity may also have resulted in emotional nuances, which are not given by Artificial Intelligences. For example, as previously indicated, there are questions about the transition from one emotion to another and about what is included in the "other" category, which we have classified with a neutral emotional label without knowing the emotional elements that define it.

Based on the combination of emotional valences, our results (Table 8), reinforce the idea of a degree of uniformity in the DeepSeek and Yiaho results, at least in one of the emotional components. On the other hand, there is very little

Table 5. Emotional couples for each participant through ChatGPT, Grok, DeepSeek and Yiaho (Prompt 1).

	ChatGPT	Grok	DeepSeek	Yiaho
I.1.	Anticipation, Sadness	Sadness, Anticipation	Sadness, Optimism, Anticipation	Optimism, Anticipation
I.2.	Sadness, Others	Sadness, Fear	Optimism, Sadness	Sadness, Optimism
I.3.	Others, Sadness	Anticipation, Anger, Happiness	Sadness, Disgust	Sadness, Disgust
I.4.	Happiness, Optimism, Anticipation	Sadness, Anticipation	Optimism, Happiness	Anticipation, Happiness
I.5.	Disgust, Anticipation	Sadness, Fear, Surprise	Sadness, Disgust	Disgust, Sadness
I.6.	Others, Optimism	Sadness, Anticipation	Optimism, Sadness	Optimism, Sadness
I.7.	Optimism, Happiness, Sadness	Sadness, Happiness	Optimism, Anticipation	Optimism, Happiness
I.8.	Surprise, Anticipation	Sadness, Fear, Anticipation	Fear, Optimism	Fear, Optimism

Table 6. Summary of emotions by frequency, Percentage, and Valence (Prompt 1).

Emotions	Nombre	Percentage	Valence
Sadness	20	0,286	-
Fear	5	0,071	-
Frustration	0	0,000	-
Anticipation	13	0,186	Neutral
Disgust	5	0,071	-
Surprise	2	0,029	Neutral
Happiness	7	0,100	+
Anger	1	0,014	-
Optimism	14	0,200	+
Others	3	0,043	Neutral
Total	70	1	

Table 7. Emotional percentages by valence category (Negative, Positive, Neutral).

Emotions	Negative emotions	Positive emotions	Neutral emotions	Total
Percentage	44,3%	30%	25,7%	1

Table 8. Valence of emotional couples through ChatGPT, Grok, DeepSeek and Yiaho (Prompt 1).

	ChatGPT	Grok	DeepSeek	Yiaho
I.1.	(neutral, -)	(-, neutral)	(-, neutral)	(+, neutral)
I.2.	(-, neutral)	(-,-)	(+,-)	(-,+)
I.3.	(neutral, -)	(neutral, neutral)	**(-,-)**	**(-,-)**
I.4.	(+,+)	(-, neutral)	(+,+)	(neutral, +)
I.5.	(-, neutral)	(-, neutral)	**(-,-)**	**(-,-)**
I.6.	(neutral, +)	(-, neutral)	**(+,-)**	**(+,-)**
I.7.	(+, neutral)	(-,+)	(+, neutral)	(+,+)
I.8.	(neutral, neutral)	(-, neutral)	**(-,+)**	**(-,+)**

correspondence between the elements of the emotional combinations generated by ChatGPT and Grok and what was provided by the other Artificial Intelligences.

In summary, from this prompt, it seems that there is a negative main emotional trend, which should, however, be put into perspective given the simultaneous presence of positive emotions, neutral ones, and the lack of uniformity in the results. In addition, we only note one observation (I.5.) for which the valence of the first component is similar for the four tools. Based on the limitations, our second prompt was designed for capturing more emotional nuances (ChatGPT's "other" category) and to reach emotional percentage sums equal to one.

Results from Prompt 2

Looking at the results (Table 9), it can be seen that the degree of precision improved significantly when moving from the first prompt to the second one. On average, more than five emotions per participant were entered into the analysis, all Artificial Intelligences included, and seven new emotional labels are put forward (see Table 9). For ChatGPT, the weight of some negative emotions that were included in the "Other" category resulted in significant changes to the participants' emotional patterns. For some data – with little influence on the results – the Artificial Intelligences sometimes highlight some emotions that are not emotions (we will call them *emotional hallucinations*); Artificial Intelligences seem

Table 9. Number of excess emotions detected by ChatGPT, Grok, DeepSeek and Yiaho (Prompt 2).

Number of excess emotions	ChatGPT	Grok	DeepSeek	Yiaho	Mean
I.1.	7	3	4	9	5,75
I.2.	14	5	3	9	7,75
I.3.	4	5	1	6	4
I.4.	3	3	2	4	3
I.5.	6	4	3	6	4,75
I.6.	5	4	2	10	5,25
I.7.	10	6	4	12	8
I.8.	12	4	4	7	6,75
Mean	7,625	4,25	2,875	7,875	5,65625

sometimes to have an *emotional imagination* that is surprising and confusing. For example, ChatGPT associates self-reflection and insouciance with emotions, and Yiaho makes the same association with relaxation (these *emotions* have been excluded from the analysis). For example, ChatGPT recognizes that self-reflection is not an emotion, which makes the result disturbing.

> *Self-reflection is not an emotion in itself, but rather a mental process. It is the process of thinking about one's own thoughts, behaviors, actions and experiences. It is a kind of inner reflection that leads to a better understanding of oneself, to the analysis of one's decisions and to greater self-awareness* (ChatGPT).

Based on this data, the second prompt provides greater granularity but requires a thorough check of the results to analyze their true relevance.

Similar to the first prompt, we work in terms of emotional couples that include the two most important emotional components for each Artificial Intelligence as well as for the lexical analysis (the first element of the couple corresponding to the most predominant emotion, see Table 10). In the appendix of the article, there is an example, for one participant, showing the results from this second prompt (see Table 15).

Table 10. Emotional couples for each participant for ChatGPT, Grok, DeepSeek, Yiaho (Prompt 2) and NRC emotion lexicon (Prompt 2).

	ChatGPT	Grok	DeepSeek	Yiaho	NRC emotion lexicon
I.1.	Regret, Anger	Sadness, Anticipation	Sadness, Anger	Frustration, Sadness	Anticipation, Sadness
I.2.	Sadness, Fear	Sadness, Fear	Sadness, Fear	Frustration, Sadness	Sadness, Anticipation
I.3.	Sadness, Disgust	Sadness, Anticipation	Sadness, Surprise	Frustration, Disgust	Sadness, Anticipation, Confidence
I.4.	Anger, Disgust	Anticipation, Anger	Sadness, Surprise	Frustration, Happiness	Anticipation, Fear, Happiness, Sadness, Confidence
I.5.	Sadness, Disgust	Sadness, Anticipation	Sadness, Disgust	Frustration, Disgust	Sadness, Anticipation
I.6.	Sadness, Disgust	Sadness Anticipation	Sadness Optimism	Uncertainty, Disgust[1], Happiness, Hope	Anticipation, Happiness
I.7.	Anger, Optimism	Sadness, Happiness	Happiness, Optimism	Optimism, Doubt	Confidence, Happiness
I.8.	Powerlessness, Anger	Sadness, Fear	Sadness, Fear	Doubt, Powerlessness	Sadness, Anticipation

[1]Same percentage for the last three emotions.

Table 11. Summary of detected emotions by frequency, Percentage, and Valence according to Prompt 2.

Emotions	Number	Percentage	Valence
Sadness	26	0,302	-
Fear	6	0,070	-
Frustration	5	0,058	-
Anticipation	12	0,140	Neutral
Disgust	8	0,093	-
Surprise	2	0,023	Neutral
Happiness	7	0,081	+
Anger	6	0,070	-
Optimism	4	0,047	+
Confidence	3	0,035	+
Powerlessness	2	0,023	-
Hope	1	0,012	+
Regret	1	0,012	-
Doubt	2	0,023	-
Uncertainty	1	0,012	-
Total	86	1	

Table 12. Emotional percentages by valence category (Negative, Positive, Neutral).

Negative emotions	Positive emotions	Neutral emotions	Total
0,663	0,174	0,163	1

Considering any emotion in the couples, sadness seems to emerge strongly (more intensely than in prompt 1, as shown by the average results in Tables 11 and 12) and, more broadly, we note a strong presence of emotions with negative dimensions which, as we had previously suggested, could be explained by a negative stock market context during the experiment. This observation is much more noticeable for the second prompt (66% for the second prompt versus 44% for the first one, see Table 12 below), which suggests that a large part of the "Other" category identified by ChatGPT mainly corresponded to negative emotions for this Artificial Intelligence. Results from the second prompt reinforce the idea of a general negative emotional charge. The transition from prompt 1 to prompt 2 did not necessarily result in a better perception of some emotional nuances; however, it did give more weight to the conclusions regarding the general direction of the emotional trend.

As we mentioned for the first prompt, it should be pointed out that the negative category may be underestimated because anticipation and surprise were considered as neutral (it can be interpreted in a positive or negative way). In fact, and this explains its second position in the couples, the results of previous transactions did not meet the participants' expectations, and so sadness prevailed over anticipation and surprise.

Comparing the results on the basis of the Artificial Intelligences hardly ever leads to homogeneous results. Table 12 shows a general trend arising - in our analysis, emotions with negative charges -, but the results are unable to uniformly capture emotional nuances and more complex elements. In addition, we also propose an analysis of couples according to the valence of emotions (Table 13).

There is no homogeneous result in any cases (contrary to the first prompt, for which some similarities between DeepSeek and Yiaho had been identified), in terms of emotional couples. However, for emotional valence, the first component is similar for all Artificial Intelligences for six out of eight observations. It should be noted that a higher degree of homogeneity could also be reached if we made the methodological choice to systematically consider anticipation as a negative emotion, which is not necessarily the case according to the written data. Furthermore, the conclusions drawn from Artificial Intelligences correspond very little from lexical analysis. Our results raise concerns. Indeed, the analysis

Table 13. **Valence of emotional couples through ChatGPT, Grok, DeepSeek and Yiaho (Prompt 2).**

	ChatGPT	Grok	DeepSeek	Yiaho	Lexical analysis
I.1.	(-,-)	(-, neutral)	(-,-)	(-,-)	(neutral,-)
I.2.	(-,-)	(-,-)	(-,-)	(-,-)	(-, neutral)
I.3.	(-,-)	(-, neutral)	(-, neutral)	(-,-)	(-, neutral)
I.4.	(-,-)	(neutral,-)	(-, neutral)	(-,+)	(neutral, neutral)
I.5.	(-,-)	(-, neutral)	(-,-)	(-,-)	(-, neutral)
I.6.	(-,-)	(-, neutral)	(-,+)	(-,+)	(neutral,+)
I.7.	(-,+)	(-,+)	(+,+)	(+,-)	(+,+)
I.8.	(-,-)	(-,-)	(-,-)	(-,-)	(-, neutral)

was conducted on short written data (more comparable to verbatim than to texts with high levels of structuring) and on a limited number of people with relatively similar sociodemographic profiles. In addition, the data was collected at the end of an experiment corresponding to the same environment for all participants.

Discussion

We answer both research questions negatively: Artificial Intelligences cannot provide homogeneous results for assessing emotional patterns in written documents. Nor does a comparison of AI results with lexical analysis validate the automatic generative method. In fact, our findings demonstrate an ability to detect a relatively stable basic emotional trend, depending on Artificial Intelligence. The finding is more striking for the most accurate prompt (prompt 2).

However, firstly, once more accuracy is required, the emotional nuances seem to be handled differently. Secondly, the complexity of the prompts does not improve our results. The most general-purpose prompt gave the best results in terms of homogeneity, at least between DeepSeek and Yiaho. On the basis of this result, which involves the building of an all-purpose emotional label (this is particularly the case for ChatGPT), conclusions in favor of the use of Artificial Intelligence could eventually be drawn, although great prudence would be required.

While we were looking for more emotional accuracy, the lack of uniformity increased significantly when more details were asked on the "Others" category. The second prompt results in the identification of emotions which were not necessarily so, as well as emotional duplications which raise questions. According to our results, there is an emotional space which is not well covered by Artificial Intelligences. The advantage of the second prompt is its ability to reinforce the hypothesis of a general negative emotional trend. In summary, the lack of uniformity in the results and the *emotional hallucinations*[50] raise the question of the relevance of using Artificial Intelligences when, for example, medical diagnoses must be made[51] or in the field of mental health.[52] We suggest that, depending on the selected generative tool, researchers could draw conclusions that best match their personality: the question arises of the researcher's reflexivity[53] and their personal emotional positioning.[29]

Thirdly, the clear definition of emotions does not always seem to be controlled and the causal links between the emotional patterns developed are not included in the analysis (for example, in the context of the study, the links between anticipation – considered as a neutral emotion, and negative emotions). What is worrying for an inexperienced user is the AI's ability to provide apparently scientific support for their answers. Fourthly, the repetition of prompts seems to ensure some results consistency within the same Artificial Intelligence, but not necessarily between them.

According to the literature, Artificial Intelligences should not be used in a purely naive and blind way – because of the substantial increase in the speed of processing information and because it provides apparently *"intelligent"* reasons for its analysis – but should be combined with other tools involving a higher degree of human commitment. In all cases, it seems necessary to stress the importance of *smart* use of automatic content generation, especially for requests where a lack of prior familiarization could be detrimental. According to us, this conclusion makes sense for systematized processing that involves highly complex elements and for which the interpretative component is an essential part of the analysis.

Concerning our work limitations, the results were drawn from a small number of observations (45 short written documents from a sample of eight students) and a very specific context (decision-making on the stock market). As the participants were paid, it was difficult to work with a larger number of people. First of all, it is clear that further research could consider a larger amount of data to be processed, as well as working on a variety of contexts in which emotions could arise. Secondly, our results could be compared with facial recognition processes. Finally, the analysis could focus on how the primary emotions identified could lead to emotional paralysis (in the case of strong negative emotions charges) or emotional euphoria (in the case of strong positive emotions charges).

Data availability

All data underlying the findings of this study are openly available on Zenodo under a Creative Commons Attribution 4.0 International license (CC-BY 4.0) license:

- The full questionnaire used in the study [1]
- Anonymized participant responses and raw data used for statistical analyses [2]
- Table 14: Results Example for Prompt 1 (Appendix 1) [3]
- Table 15: Results Example for Prompt 2 (Appendix 2) [4]
- The consent form for publication [5]

These materials support the reproducibility of the study and illustrate the qualitative analysis procedure.

1. Lexico-discursive Analysis of Emotions. Zenodo. https://doi.org/10.5281/zenodo.15374312[54]
2. Questionnaire responses [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15374745[55]
3. Table 14 Appendix 1 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15376528[56]
4. Finet, A., Kristoforidis, K., & Laznicka, J. (2025). Table 15 Appendix 2 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15376535[57]
5. Consent Form for Publication. Zenodo. https://doi.org/10.5281/zenodo.15387407[58]

References

1. Khare SK, Blanes-Vidal V, Nadimi ES, *et al.*: **Motion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations.** *Inf. Fusion.* 2024; **102**: 102019.
Publisher Full Text
2. Luan H, Geczy P, Lai H, *et al.*: **Challenges and Future Directions of Big Data and Artificial Intelligence in Education.** *Front. Psychol.* 2020; **11**: 580820.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
3. Vistorte AOR, Deroncele-Acosta A, Ayala JLM, *et al.*: **Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review.** *Front. Psychol.* 2024; **15**: 1387089.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
4. Hasan ME, Islam MJ, Islam MR, *et al.*: **Applications of Artificial Intelligence for Health Informatics: A Systematic Review.** *Journal of Artificial Intelligence and Machine Sciences.* 2023; **4**(2): 19–46.
Publisher Full Text
5. Al Bashar M, Khan IH: **Artificial Intelligence in Industrial Engineering A Review.** *Int. J. Sci. Eng. Res.* 2017; **2**(3): 9.
6. Plathottam SJ, Rzonca A, Lakhnori R, *et al.*: **A review of artificial intelligence applications in manufacturing operations.** *Journal of Advanced Manufacturing and Processing.* 2023; **5**(3): e10159.
Publisher Full Text
7. Schwab A, Shuumarjav Y, Telkamp JB, *et al.*: **Artificial Intelligence and the Operationalization of Psychological Constructs: The Case of Emotions and Emotional Authenticity of Entrepreneurs During Pitch Presentations.** *Methods to Improve Our Field Research Methodology in Strategy and Management.* Emerald Publishing Limited; 2023; vol. **14**: 133–155.
Publisher Full Text
8. Thakur P, Shrivastava R: **A review on text based emotion recognition system.** *International Journal of Advanced Trends in Computer Science and Engineering.* 2018; **7**(5): 67–71.
Publisher Full Text
9. Raval AS, Lohia A: **A Survey on Techniques, Methods, and Applications of Text Analysis.** *IJCRT.* 2021; **9**(5): 181–184.
10. Chutia T, Baruah N: **A review on emotion detection by using deep learning techniques.** *Artif. Intell. Rev.* 2024; **57**: 203–282.
Publisher Full Text
11. Machová K, Szabóova M, Paralič J, *et al.*: **Detection of emotion by text analysis using machine learning.** *Front. Psychol.* 2023; **14**.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
12. Bharti SK, Varadhaganapathy S, Gupta RK, *et al.*: **Research Article Text-Based Emotion Recognition Using Deep Learning Approach.** *Comput. Intell. Neurosci.* 2022; **2022**: 8 pages. Article ID 2645381.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
13. Ahamad R, Mishra KN: **Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach.** *J. Big Data.* 2025; **12**(11).
Publisher Full Text
14. Murthy AR, Kumar AKM: **A Review of Different Approaches for Detecting Emotion from Text.** *IOP Conference Series: Materials Science and Engineering.* 2021; **1110**(1): 012009.
Publisher Full Text

15. Bryda G, Sadowski D: **From Words to Themes: AI-Powered Qualitative Data Coding and Analysis in Computer Supported Qualitative Research.** *New Trends in Qualitative Research (WCQR2024).* Springer; 2024; 309–345.
Publisher Full Text

16. Mohammad SM, Turney PD: **Crowdsourcing a Word–Emotion Association Lexicon.** *Comput. Intell.* 2013; **29**(3): 436–465.
Publisher Full Text

17. Zhang H, Wu C, Xie J, *et al.*: **Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis.** 2024.
Publisher Full Text

18. Morgan DL: **Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT.** *Int. J. Qual. Methods.* 2023; **22**: 1–10.
Publisher Full Text

19. Salah M, Al Halbusi H, Abdelfattah F: **May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research.** *Comput. Hum. Behav. Artif. Hum.* 2023; **1**: 100006.
Publisher Full Text

20. Özden M: **Content and thematic analysis techniques in qualitative research: Purpose, process and features.** *Qualitative Inquiry in Education: Theory & Practice.* 2024; **2**(1): 64–81.
Publisher Full Text

21. Jalali M: **Integrating AI language models in qualitative research: Replicating interview data analysis for creating causal loop diagrams.** *Syst. Dyn. Rev.* 2024; **40**.
Publisher Full Text

22. Deng J, Ren F: **A Survey of Textual Emotion Recognition and Its Challenges.** *IEEE Trans. Affect. Comput.* 2021; **14**(4): 49–67.
Publisher Full Text

23. Vicci H: **Emotional Intelligence in Artificial Intelligence: A Review and Evaluation Study.** *SSRN Electron. J.* 2024, May.
Publisher Full Text

24. Hitch D: **Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future?** *Qual. Health Res.* 2024; **34**(7): 595–606.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Grimmer J, Stewart BM: **Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.** *Polit. Anal.* 2013; **21**(3): 267–297.
Publisher Full Text

26. Sinha R, Li Y, Jona K, *et al.*: **The Role of GenAI in Qualitative Analysis: Grounded Theory Approach.** *Learning with Large Language Models for Education.* 2024.
Publisher Full Text

27. Christou PA: **How to Use Artificial Intelligence (AI) as a Resource, Methodological and Analysis Tool in Qualitative Research?** *Qual. Rep.* 2023; **28**(7): 1968–1980.
Publisher Full Text

28. Pattyn F: **The Value of Generative AI for Qualitative Research: A Pilot Study.** *J. Data Sci. Intell. Syst.* 2024; 1–10.
Publisher Full Text

29. Lustick H, Yang X, Hakouz A: **The Role of Emotions in Qualitative Analysis: Researchers' Perspectives.** *Qual. Rep.* 2024; **29**(4): 1103–1124.
Publisher Full Text

30. Ackert LF, Church BK, Tompkins J, *et al.*: **What's in a name? An experimental examination of investment behavior.** *Rev. Financ.* 2005; **9**(2): 281–304.
Publisher Full Text

31. Biais B, Hilton D, Mazurier K, *et al.*: **Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market.** *Rev. Econ. Stud.* 2005; **72**(2): 287–312.
Publisher Full Text

32. Bruguier AJ, Quartz SR, Bossaerts P: **Exploring the nature of "trader intuition".** *J. Financ.* 2010; **65**(5): 1703–1723.
Publisher Full Text

33. Widyarini I: **The role of negative moral emotions (anger and disgust) in ethical decision making.** *8th International Conference of Asian Association of Indigenous and Cultural Psychology (ICAAIP 2017).* Atlantis Press; 2017; pp. 244–250.
Publisher Full Text

34. Dorn D, Sengmueller P: **Trading as Entertainment.** *Manag. Sci.* 2009; **55**(4): 591–603.
Publisher Full Text

35. Finet A, Kristoforidis K, Viseur R: **L'émergence de biais comportementaux en situation de trading: une étude exploratoire.** *Revue Recherches en Sciences de Gestion.* 2021; **146**: 147–182.
Publisher Full Text

36. Finet A, Laznicka J: **Addressing Emotional Dysregulation in Experimental Design.** *Psychology and Behavioral Sciences.* 2025; **14**(1): 7–18.
Publisher Full Text

37. Grether D: **Financial Incentive Effects and Individual Decision-making.** *Social Science Working Papern, n°401.* 1981; p. 13.

38. Ricciardi R: *A Risk Perception Primer: A Narrative Research Review of the Risk Perception Literature in Behavioral Accounting and Behavioral Finance.* 2004; p. 109.
Reference Source

39. Bloomfield R, Anderson A: **Experimental Finance.** *Johnson School Research Paper Series, n°23.* 2009; p. 22.
Publisher Full Text

40. Libby R, Bloomfield R, Nelson MW: **Experimental research in financial accounting.** *Acc. Organ. Soc.* 2002; **27**(8): 775–810.
Publisher Full Text

41. Domeignoz C, Morin E: **Les émotions ont leur raison, qu'il convient d'écouter.** *Entreprendre & Innover.* 2016; **2**: 7–15.
Publisher Full Text

42. Waddington K: **Using qualitative diary research to understand emotion at work.** *A Day in the Life of a Happy Worker.* Psychology Press; 2012; pp. 132–149.

43. Scott S: **'I enjoy having someone to rant to, I feel like someone is listening to me': Exploring emotion in the use of qualitative, longitudinal diary-based methods.** *Int. J. Qual. Methods.* 2022; **21**.
Publisher Full Text

44. Nandwani P, Verma R: **A review on sentiment analysis and emotion detection from text.** *Soc. Netw. Anal. Min.* 2021; **11**(1): 81.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Pradhan A, Senapati MR, Sahu PK: **Comparative analysis of lexicon-based emotion recognition of text.** *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021.* Singapore: Springer Nature Singapore; 2023; pp. 671–677.
Publisher Full Text

46. Ekman P: **Are there basic emotions?** *Psychol. Rev.* 1992; **99**(3): 550–553.
Publisher Full Text

47. Plutchik R: **A General Psychoevolutionary Theory of Emotion.** *Emotion: Theory, research, and experience.* 1980; **1**(3): 3–33.
Publisher Full Text

48. Polo EM, Farabbi A, Mollura M, *et al.*: **Understanding the role of emotion in decision making process: using machine learning to analyze physiological responses to visual, auditory, and combined stimulation.** *Front. Hum. Neurosci.* 2024; **17**.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Mohammad SM, Turney PD: **Nrc Emotion Lexicon.** *National Research Council, Canada.* 2013; **2**: 234.

50. Azamfirei R, Kudchadkar SR, Fackler J: **Large language models and the perils of their hallucinations.** *Crit. Care.* 2023; **27**: 120.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Mohana Sundaram AS, Patil B, Praticò D: **ChatGPT's Inconsistency in the Diagnosis of Alzheimer's Disease.** *J. Alzheimer's Dis. Rep.* 2024; **8**: 923–925.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Wake N, Kanehira A, Sasabuchi K, *et al.*: **Bias in Emotion Recognition with ChatGPT.** 2023. arXiv preprint arXiv: 2310.11753v2.
Publisher Full Text

53. Ide Y, Beddoe L: **Challenging perspectives: Reflexivity as a critical approach to qualitative social work research.** *Qual. Soc. Work.* 2024; **23**(4): 725–740.
Publisher Full Text

54. Finet A, Kristoforidis K, Laznicka J: **Lexico-discursive Analysis of Emotions.** *Zenodo.* 2025.
Publisher Full Text

55. Finet A, Kristoforidis K, Laznicka J: Questionnaire responses. [Data set]. *Zenodo.* 2025.
Publisher Full Text

56. Finet A, Krisoforidis K, Laznicka J: Table 14 Appendix 1. [Data set]. *Zenodo.* 2025.
Publisher Full Text

57. Finet A, Kristoforidis K, Laznicka J: Table 15 Appendix 2. [Data set]. *Zenodo.* 2025.
Publisher Full Text

58. Finet A, Kristoforidis K, Laznicka J: **Consent Form for Publication.** *Zenodo.* 2025.
Publisher Full Text