Context over Categories: Implementing the Theory of Constructed Emotion with LLM-Guided User Analysis

ANONYMOUS AUTHOR(S)

Emotion analysis is a critical research area with applications in content moderation and personalized systems. Many existing approaches rely on Ekman's universal emotions theory, which reduces emotions to static categories, neglecting their complexity and contextual variability. This work introduces a novel, context-aware approach based on Lisa Feldman Barrett's Theory of Constructed Emotion. A key contribution is the development of the "context sphere," a personalized construct derived from user behavior data. To our knowledge, this is the first operationalization for computational methods. A context-aware emotion analysis pipeline was developed, incorporating an advanced Large Language Model (LLM) prompting strategies like role-play and controlled generation. A case study in content moderation demonstrates how the "context sphere" enables contextually aware emotion analyses. Future directions include refining the framework, advancing LLM methodologies, and conducting user studies. This research lays the foundation for more human-centered, ethical, and effective emotion analysis systems.

CCS Concepts: • **Computing methodologies** → *Natural language generation*; • **Human-centered computing** → **HCI theory, concepts and models**.

Additional Key Words and Phrases: Human-Computer Interaction, Emotion Analysis, Large Language Models (LLMs), Context Aware Computing, Online Content Moderation

ACM Reference Format:
Anonymous Author(s). 2025. Context over Categories: Implementing the Theory of Constructed Emotion with LLM-Guided User Analysis. 1, 1 (January 2025), 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Emotion analysis has become an increasingly significant area of research, with applications spanning sentiment analysis, content moderation, and human-centered adaptation. Despite its prevalence, much of the existing work in this domain continues to rely on traditional models, such as Paul Ekman's theory of universal emotions [1, 16]. This framework, while widely adopted, simplifies the complexity of human emotion into a set of discrete categories. Its simplicity has also shaped recent work in informatics, where emotion analysis often relies on predefined categories, particularly in natural language processing (NLP) applications such as social media analysis. These approaches frequently focus on individual words, sentences, and basic sentiments [3, 45]. Moreover, resources like the NRC Word-Emotion Association Lexicon exemplify this trend by relying on fixed emotion labels to associate textual data with emotions [1, 32]. However, such methods, rooted in Ekman's framework, are inherently limited in their ability to account for the nuanced and context-dependent nature of human emotions. These oversimplifications fail to capture the complexities of emotional expression and perception [34]. For instance, as Kate Crawford highlights *Atlas of AI* [14], such simplifications pose significant societal and ethical risks. Overreliance on rigid models, like Ekman's, can reinforce biases, such as racial profiling and stereotypes, by reducing human emotions to fixed categories and predefined responses. A more nuanced

Manuscript submitted to ACM

approach is essential to avoid these pitfalls and to reflect the breadth, diversity, and complexity of emotional expression. Addressing these concerns is critical for ensuring that advancements in emotion analysis technologies are not only accurate but also equitable and ethically sound.

The work by highly influential cognitive scientist Lisa Feldman Barrett offers a state-of-the-art alternative, arguing that emotions are not innate and universally recognized but are constructed through individual experiences and contextual factors [5, 7, 9]. Barrett's Theory of Constructed Emotion contrasts with Paul Ekman's concept of universal emotions by challenging reducing emotions to simplistic abstractions. Instead, it recognizes the inherently complex and constructed nature of emotional experiences. Although informatics often build upon simplified abstractions using machine learning models and other classifiers, the advancements in Large Language Models (LLMs) offer a significant opportunity to revisit this paradigm. This highlights a research gap: How can we leverage the capabilities of LLMs to move beyond a typological view of emotions and incorporate their complex, constructed nature into an information system?

This Late-Breaking Work Submission aims to close this gap by showing a novel approach using advanced LLM guidance techniques to operationalize Barrett's Theory of Constructed Emotion. Specifically, considerations are made regarding how data needs to be preprocessed and how LLMs can be guided to align with this psychological concept. The primary contribution is a conceptual research framework (Figure 1) that outlines the methodological considerations and inherent limitations of such a system. The Design Science Research framework [25, 26] is used to develop and design an artifact, which represents an instance of this framework. Through this instance, the potential for nuanced emotion understanding in textual data is explored. This is illustrated through the case of online content moderation, where large amounts of data need to be checked, often making it infeasible for humans to read and conduct a condensed and nuanced user analysis. To achieve this, state-of-the-art LLM guidance techniques are used, such as role-play prompting [24, 28], controlled generation [19] and meta-prompting [46]. Evaluation is separated into using an LLM as a judge approach [29, 47]. Future research steps will integrate a user study.

Contributions. This Late-Breaking Work submission makes the following contributions to the field of human-computer interaction with a focus on advancing emotion analysis:

- **Novel Methodological Framework:** A context-aware emotion analysis framework that integrates advanced LLM guidance techniques to operationalize emotions in a nuanced and dynamic way.
- **Innovation in Emotion Modeling:** Introduction of the "context sphere" as a data-driven construct for modeling emotions in real-world applications, such as online content moderation.
- **Application Potential:** Demonstrates how LLMs can be adapted for context-sensitive emotion analysis in challenging domains, addressing both technical and practical limitations of existing methods.
- **Provocation for Future Work:** Offers a foundation for further exploration of context-aware computational methods and their alignment with complex emotional phenomena, sparking novel research conversations within HCI.

These contributions collectively advance the state of the art in emotion analysis by bridging the gap between cognitive science and computational systems, enabling more nuanced, context-aware applications in areas such as sentiment analysis, personalization, and content moderation.

2 Related Work

Recent emotion classification in informatics heavily relies on Ekman's universal emotions theory, categorizing emotions into basic, universally recognized types [1, 16] which was originally developed over 50 years ago. Its simplicity seems to still appeal to researchers, influencing algorithm development and in NLP, especially in social media analyses [3, 45]. Supervised learning approaches using labeled datasets prefer simplified classifications, limiting recognition of implicit emotions and expression complexities [1, 2, 40]. Lexicons like the NRC Word-Emotion Association Lexicon map words to basic emotion classifications [32], reflecting a simplified, categorical approach to emotion representation [1]. Critiques highlight this approach's oversimplification, overlooking the nuanced, constructed nature of emotions emphasized by Lisa Feldman Barrett [5, 7]. Barrett argues that emotions are contextually constructed and influenced by individual and cultural differences. Studies demonstrate cultural variability in emotion perception [18] and language's influence on emotion [8], challenging Ekman's theory on the universal emotion theory.

Moreover, there are calls to bridge the gap between the advancements in cognitive science and emotion analysis to better capture the complex nature of emotions [34]. Current emotion recognition practices often rely on rule-based and learning-based approaches focusing on Ekman's categories [1]. More recently, deep learning models have introduced complexity. Nevertheless, they classify emotions into discrete categories [1]. Hybrid approaches integrating multimodal data offer insights into addressing these challenges [38]. Advancements in contextualized language models like BERT improve emotion recognition by capturing nuanced expressions [15], yet do not fully embrace Barrett's model accounting for variability and contextuality in emotions [31]. Recognizing these limitations underscores the need to integrate flexible, context-sensitive models into emotion recognition systems [1].

3 Methodology

The primary research outcome is the creation of a context-aware emotion analysis pipeline using Large Language Models (LLMs). This work consists of three parts: the preprocessing in building the "context sphere", the LLM-Pipeline, and the evaluation of the intermediate steps.

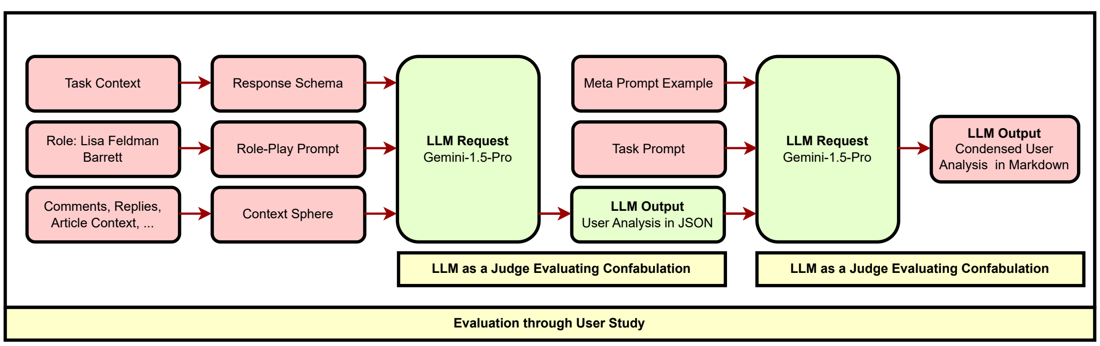

Fig. 1. Pipeline from Preprocessing to Final Output - Please note: User Study not part of this Late-Breaking Work Submission.

3.1 Context Sphere Building

The preprocessing is the starting point of the conceptual research framework seen in Figure 1, crucial for transforming raw data into a usable format for subsequent LLM usage. Our data originates from "(anonymised)," the online portal of a national newspaper, and focuses specifically on the comment sections and related articles. The dataset contains publication details, comment content, and related articles over a 30-day period involving 23,925 users in May 2019. The primary goal is to create a document that encapsulates essential situational context for each user. The term "context sphere" is chosen to reflect the core of Lisa Barrett's theory that emotions arise from a complex interplay of varied and interrelated experiences and contexts [7]. Much like a sphere uniformly encloses space in three dimensions, our "context sphere" gathers data to view a user's interactions from multiple angles, allowing for nuanced emotional interpretations influenced by various contextual layers. In this concrete case, this involves including all comments a user has made in May 2019, along with the surrounding context. This context comprises article metadata, a description of the "context sphere", and interactions between the user and others. When a user replies to a conversation, the entire thread, including comments from other participants, is added into the context sphere up to a designated cutoff point. This cutoff point is a condition which becomes true if a comment from the analyzed user is the last comment inside a comment thread. This means that the user can engage in an extended discussion, and every participating comment — whether from the analyzed user or others — will be included in the "context sphere." The cutoff point is used, since the analysis should focus on the user's context, capturing every interaction up to their last known comment. This approach not only reduces the overall length of the conversation being analyzed but also centers the analysis on the user's perspective. The resulting document termed the "context sphere," includes the user's contributions and the surrounding context, such as the fact that it originates from a national online newspaper, the time frame, and other supplemental information. This differs significantly from many traditional methods in informatics and psychology that rely on keyword or single sentences [1, 32, 33], which often overlook the crucial role of context emphasized by Barrett [5, 7, 9].

To maintain privacy and reduce biases linked to gender stereotypes, personal identifiers like usernames and gender are excluded from the context sphere. In line with Barrett's emphasis on context, our context sphere captures complete interaction footprints within discussions. A key decision in the methodology was to apply a selective pruning process with the cutoff point rather than including entire threads with potentially hundreds of comments and sub-threads. By balancing the need for comprehensive contextual data with practical considerations of data efficiency and system limitations, we ensure that the analysis remains both robust and manageable, allowing for meaningful insights without overwhelming the system or compromising user privacy.

The chosen format of the context sphere is Markdown, prioritizing readability for both humans and LLMs while maintaining the thread structure of conversations and minimizing token use. Although XML and JSON are commonly recommended for structuring prompts [22, 23, 36], they come with drawbacks in this scenario. XML, while useful for distinct data blocks, adds redundant tags in repetitive structures like comment threads, leading to inefficiency in token usage. JSON lacks easy readability, which is essential for development but is also considered a valid choice. Markdown is selected as it fulfills our requirements by balancing clarity and efficiency, facilitating easier navigation for both human evaluators and LLMs during analysis.

3.2 LLM Guidance

This research introduces a novel approach combining the Theory of Constructed Emotion with advanced LLM guidance techniques. A core challenge is classifying online user behavior using Barrett's theory, specifically avoiding predefined emotion categories and fixed identifiers [6, 7, 10]. Recognizing the variability of emotional expression, our system analyses the context and uses its exceptional language capabilities to describe the human emotional landscape, distinguishing it from traditional methods using predefined categories [1, 45]. This constructionist view informs our preprocessing and the LLM pipeline (Figure 1), where the LLM, in the role of Barrett, performs the user analysis. Key advancements are enabled by the rapid development of LLM with the increased size of context windows, enhanced reasoning capabilities, sophisticated role-play prompting, and controlled output generation [19–21, 37].

Both constructed emotions and LLM outputs are inherently probabilistic and context-dependent, contrasting with fixed views of emotions and deterministic LLMs. This shared probabilistic nature presents a research challenge due to non-deterministic outcomes and the absence of definitive ground truth, aligning with the "population thinking" of the Theory of Constructed Emotion. This paper presents an approach to this challenge, using only some of the possible techniques in preprocessing and LLM guidance.

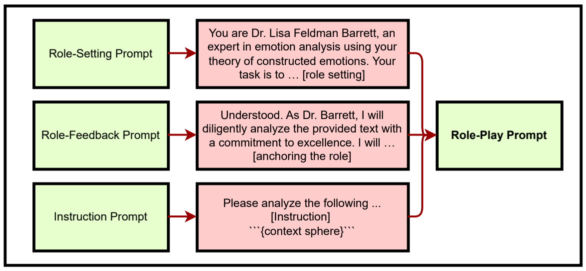

Fig. 2. Role-Play Prompting according to the example from & Kong et al.. [28].

The system utilizes a context sphere (Markdown) within a role-play prompt, which lets the LLM impersonate Barrett (Figure 2), a technique shown to enhance reasoning [28] and shown to be feasible [24]. Before the API request to Gemini-1.5 Pro (LLM), a JSON schema enforces controlled generation, defining a multi-class structure with five main fields. The (1) "Core Affect Analysis" field, aligning with Barrett's suggestions [4, p.30], includes valence (good/bad) and arousal (activated/deactivated) as sub-fields. (2) "Cognitive Appraisal and Conceptualization" reflects Barrett's view on the role of cognition and past experiences [4, p.21]. The (3) "Cultural and Social Context" field recognizes that cultures transmit emotional meanings [7, p.910] and social contexts influence emotions [7, p.909], drawing on the constructionist perspective of emotions shaped by experience, context, culture, and language [4, 7, 39]. The (4) "Emotion Construction Analysis" field forces the LLM to combine the factors from the previous three fields, following a chain-of-thought approach [43] and using the autoregressiveness of LLM. Finally, the field (5) "Emotional Dynamics and Changes" captures the dynamic nature of emotions, contrasting with static views, and leverages the context sphere and the already generated part of the response to describe emotional dynamics without fixed entities. Each of these five main fields contains the sub-fields: (a) "Thought Process," detailing the LLM's reasoning; (b) "Analysis," presenting the classification; (c) "Observable Patterns"; (d) "Observable Anomalous Behavior"; and (e) "Rationale," guiding the LLM through the generation of the response. This type of guidance contributes to further immersing the LLM in its role. The only exception is (1), where instead of (b) "Analysis" fields for "Arousal" and "Valence" are inserted. The full LLM request consists of a role-setting prompt, role-feedback prompt, the user task prompt with the "context sphere", and the outlined response schema consisting of main classes (1-5) and sub-fields (a-e). The LLM, in this case Gemini-1.5-Pro, is only able to answer in a structured JSON output. To enhance readability and provide a usable document describing the analyzed user, the output of the first LLM request, is inserted into a second request with a meta-prompt [46]. This meta prompt instructs the model to generate a condensed Markdown report, which contains the major insights, providing a compromise that is in its length readable but still provides nuanced insights into a user online landscape. This two-step process yields a structured analysis for evaluation and a human-readable report.

3.3 Evaluation

Evaluating the LLM pipeline presents significant challenges due to the combination of the Theory of Constructed Emotion and the large volume of processed text. As depicted in Figure 1, an initial evaluation occurs after the first LLM request, serving as a crucial point for iterative improvements to prompts and outputs. This evaluation leverages LLMs as judges, a methodology supported by research demonstrating strong agreement between LLM and human evaluation[13, 47]. This approach is necessary due to the substantial size of preprocessed "context spheres" ranging from 100 up to 100,000 tokens. The efficiency of LLM as judges allows for the thorough processing of this volume of text, a capability infeasible for human evaluators within similar time constraints. Furthermore, employing LLMs for continuous feedback is a common practice in LLM pipeline development, enabling the rapid identification of potential errors and areas for refinement. This immediate feedback loop allows for rapid iterative development. A specific focus of this evaluation is the detection of confabulations commonly known as hallucinations – confident yet misleading outputs [11] – which this method is designed to mitigate. Our approach involves confabulations checks by GPT-4o, Claude 3.5 Haiku/Sonnet, and Gemini 1.5 Flash. All models receive the same task, which is the check for confabulation in the output, based on the provided input. The prompt used is inspired by Zheng et al., and enriched the redefined term of "confabulation" [11]. While numerous hallucination evaluation methods exist, many require a ground truth that we do not have[17, 27, 30], making our chosen method a pragmatic solution for our specific case. While LLMs are used as judges for evaluation in this case, further research into the robustness and general applicability of these methods is needed. The user study results will form part of the mature study and future publications.

4 Results and Analysis

The primary goal of this research is to introduce a novel approach to emotion analysis by operationalizing the Theory of Constructed Emotion, representing a significant departure from traditional emotion classification methods commonly used in psychology and informatics. A key limitation lies in the absence of a definitive ground truth for emotions, as emotions are individually constructed and context-dependent [7, 10]. Consequently, external emotion analysis, including this approach, inherently involves some level of abstraction and approximation. This challenge also applies to LLMs, which rely on self-attention mechanisms to abstract and weight contextual relationships rather than storing exact information [42]. Similar to the way the brain constructs emotions, LLM outputs are inherently predictive, requiring careful interpretation to ensure accurate and meaningful insights.

In Table 1 five snippets from LLM request resulting in the user analysis in JSON, according to Figure 1 is shown. The table fields are according to the JSON data fields (1-5) and (a-e), described in the LLM Guidance Part. For each class in column one, there is one example picked from a sub-field from column two. The example call chosen had an overall of 21,671 tokens, cost 0.0353425$, and took 60.50 seconds to complete. Reading the full input and output would have taken an average reader about 63 minutes [12].

The *first example* shows the *arousal* subfield from the *Core Affect Analysis*, which gives insights about the emotional intensity but is even more specific about the complex emotional states and their interplay. The *second example*, from *Cognitive Appraisal & Conceptualization*, highlights the LLM's capability of analyzing the user's interpretive lens, identifying a pattern of skepticism and cynicism, particularly towards differing political viewpoints. This aligns with Barrett's theory by considering the cognitive processes influencing emotional expression. The *third example* in the *Cultural & Social Context* field demonstrates the LLM's initial *Thought Process*, which contextualizes the user's expressions within the specific platform ("anonymized"), the Austrian political climate during the data collection period,

Table 1. Examples output from first LLM request

Field	Sub-Field	Output
Core Affect Analysis	Arousal	Generally high, fluctuating between agitated and calmly contemptuous. [...]
Cognitive Appraisal & Conceptualization	Analysis	[...] "The user's interpretations are often filtered through a lens of skepticism and cynicism, particularly towards opposing political views. [...]
Cultural & Social Context	Thought Process	I will examine the cultural and social context by considering the platform ("anonymized"), the political climate of Austria in May 2019 (pre-election period), and the specific topics discussed (immigration, politics, media). [...]
Emotion Construction Analysis	Rationale	[...] Their negative emotional responses are often amplified by their pre-existing beliefs and the context of the online forum. The occasional deviations from this pattern [...]
Emotional Dynamics & Changes	Anomalous Observations	The user's positive comment about Fendrich deviates from their usual negative pattern, suggesting a momentary shift in emotional state. [...]

and the topics discussed. This suggests that the LLM obeys its role and considers the surrounding environment and context, which shapes the user's emotional construction.

In *Emotion Construction Analysis*, the *Rationale* example shows how the LLM synthesizes previous observations, explaining how pre-existing beliefs and the online forum context amplify negative emotional responses. It also acknowledges deviations from this pattern, hinting at the dynamic nature of emotions.

Finally, the *Emotional Dynamics & Changes* example points out an *Anomalous Observation*, where a positive comment deviates from the user's usual rather negative pattern. This highlights the potential for capturing shifts in emotional state over time and within different contexts, showcasing the fluidity of emotions.

The examples show a practical implementation of a complex psychological concept, the Theory of Constructed Emotion. The output shows that the model is aware of its role, task, and context it should work with. The proposed approach shows an alternative way of understanding a person's online behavior. Moreover, the output does not rely on simplified or predefined classification labels but makes use of the reasoning and parametric knowledge of the LLM. This work addresses the critique on typological emotion concepts and shows an alternative possible way to analyze emotions.

4.1 Discussion

Conclusion. This work advances the state of the art in emotion analysis by bridging the gap between cognitive science and computational systems, paving the way for more nuanced, context-aware applications in areas such as sentiment analysis, personalization, and content moderation. By operationalizing Barrett's Theory of Constructed Emotion through the development of the "context sphere" and employing advanced LLM guidance techniques like role-play prompting and controlled generation, this study demonstrates how dynamic and detailed emotion insights can be achieved. As a Late-Breaking Work submission, this study also highlights avenues for future exploration. These include refining the proposed framework, addressing its inherent challenges, and conducting planned user studies to validate and extend its practical applicability.

The advancements presented in this submission have significant implications for the development of more accurate and effective human-centered technologies. By operationalizing Barrett's Theory of Constructed Emotions [5] through an LLM-based approach, this work captures the complexity and context of human emotions far more effectively than models rooted in Ekman's assumptions of universal emotions. The improvements will be found in reduced false positives and a deeper contextual understanding—represent a critical step forward. To illustrate, in content moderation, this approach will help to avoid the oversimplified analysis of isolated words or small text chunks, which today can lead to unjustified actions, such as blocking users based on stylistic expressions. Instead, it will allow for a nuanced interpretation of user behavior, resulting in consequences that are better grounded, less biased, and more sensitive to context. These contributions address the ethical risks highlighted by Kate Crawford [14], including the dangers of racial profiling and algorithmic bias stemming from oversimplified models of emotion. Furthermore, this work aligns with responsible AI initiatives, such as Digital Humanism [44], which emphasize the importance of ethical, human-centered approaches in technology design. By advancing the translation of complex psychological theories into applied computational frameworks, this research lays the foundation for future work on adaptive, transparent, and context-aware systems that respect the diversity and complexity of human emotional experiences.

Limitations and Future Work. This work faces limitations due to the dynamic nature of how emotions are defined in the Theory of Constructed Emotion as well as in the constraints of current LLMs. Translating Barrett's framework into a practical system required iterative development to balance flexibility with structure. The "context sphere" captures user-specific context while minimizing complexity, but achieving consistent precision with generalized LLMs like ChatGPT remains challenging. Techniques like role-play prompting, where the LLM impersonates Barrett, provided a partial solution but rely on the assumption that the LLM's parametric knowledge aligns with her framework. The lack of existing computational applications of the Theory of Constructed Emotion and the absence of a ground truth for emotions complicate standard accuracy assessments, highlighting the need for new evaluation frameworks tailored to this theory.

Future work should address these challenges by exploring where transitions from universal emotion theory to this more dynamic model are necessary and beneficial. Human evaluation, while infeasible at scale, remains critical [35], and tailored evaluation criteria must be developed for specific tasks [41]. As results become more dynamic, the evaluation must also become more dynamic and reflect more nuanced criteria beyond accuracy. In sum, these contributions offer a foundation for future research and provoke new conversations about the integration of cognitive science into computational emotion analysis.

References

[1] Nourah Alswaidan and Mohamed El Bachir Menai. 2020. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems* 62, 8 (Aug. 2020), 2937–2987. doi:10.1007/s10115-020-01449-0

[2] Saima Aman and Stan Szpakowicz. 2007. Identifying Expressions of Emotion in Text. In *Text, Speech and Dialogue*, Václav Matoušek and Pavel Mautner (Eds.). Springer, Berlin, Heidelberg, 196–205. doi:10.1007/978-3-540-74628-7_27

[3] Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. doi:10.48550/arXiv.2010.12421 arXiv:2010.12421 [cs].

[4] Lisa Feldman Barrett. 2006. Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review* 10, 1 (Feb. 2006), 20–46. doi:10.1207/s15327957pspr1001_2

[5] Lisa Feldman Barrett. 2017. *How Emotions Are Made: The Secret Life of the Brain.* Pan Macmillan. Google-Books-ID: vjIvDQAAQBAJ.

[6] Lisa Feldman Barrett. 2017. The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience* 12, 1 (2017), 1–23. Publisher: Oxford University Press.

[7] Lisa Feldman Barrett. 2022. Context reconsidered: Complex signal ensembles, relational meaning, and population thinking in psychological science. *American Psychologist* 77, 8 (Nov. 2022), 894–920. doi:10.1037/amp0001054

[8] Lisa Feldman Barrett, Kristen A. Lindquist, and Maria Gendron. 2007. Language as context for the perception of emotion. *Trends in Cognitive Sciences* 11, 8 (Aug. 2007), 327–332. doi:10.1016/j.tics.2007.06.003 Publisher: Elsevier.

[9] Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. 2011. Context in Emotion Perception. *Current Directions in Psychological Science* 20, 5 (Oct. 2011), 286–290. doi:10.1177/0963721411422522 Number: 5 Publisher: SAGE Publications Inc.

[10] Lisa Feldman Barrett and Christiana Westlin. 2021. Navigating the science of emotion. In *Emotion measurement*. Elsevier, 39–84.

[11] Elijah Berberette, Jack Hutchins, and Amir Sadovnik. 2024. Redefining "Hallucination" in LLMs: Towards a psychology-informed framework for mitigating misinformation. doi:10.48550/arXiv.2402.01769 arXiv:2402.01769 [cs] version: 1.

[12] Marc Brysbaert. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language* 109 (Dec. 2019), 104047. doi:10.1016/j.jml.2019.104047

[13] Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. doi:10.48550/arXiv.2403.04132 arXiv:2403.04132 [cs].

[14] Kate Crawford. 2021. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. doi:10.18653/v1/N19-1423

[16] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124–129. doi:10.1037/h0030377 Place: US Publisher: American Psychological Association.

[17] Jinlan Fu, See-Kiong Ng, Zhengbao Jiang, and Pengfei Liu. 2024. GPTScore: Evaluate as You Desire. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 6556–6576. doi:10.18653/v1/2024.naacl-long.365

[18] Maria Gendron, Debi Roberson, Jacoba Marieta Van Der Vyver, and Lisa Feldman Barrett. 2014. Cultural Relativity in Perceiving Emotion From Vocalizations. *Psychological Science* 25, 4 (April 2014), 911–920. doi:10.1177/0956797613517239

[19] Google. 2024. Controlled generation | Generative AI on Vertex AI. https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/control-generated-output

[20] Google. 2024. Our next-generation model: Gemini 1.5. https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/

[21] Google. 2025. Long context | Generative AI on Vertex AI. https://cloud.google.com/vertex-ai/generative-ai/docs/long-context

[22] Google. 2025. Structure prompts | Generative AI on Vertex AI. https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/structure-prompts

[23] Jia He, Mukund Rungta, David Koleczek, Arshdeep Sekhon, Franklin X. Wang, and Sadid Hasan. 2024. Does Prompt Formatting Have Any Impact on LLM Performance? doi:10.48550/arXiv.2411.10541 arXiv:2411.10541 [cs].

[24] Steffen Herbold, Alexander Trautsch, Zlata Kikteva, and Annette Hautli-Janisz. 2024. Large Language Models can impersonate politicians and other public figures. doi:10.48550/arXiv.2407.12855 arXiv:2407.12855 [cs].

[25] Alan R Hevner. 2007. A three cycle view of design science research. *Scandinavian journal of information systems* 19, 2 (2007), 4.

[26] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS quarterly* (2004), 75–105. Publisher: JSTOR.

[27] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12 (March 2023), 248:1–248:38. doi:10.1145/3571730

[28] Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong. 2024. Better Zero-Shot Reasoning with Role-Play Prompting. doi:10.48550/arXiv.2308.07702 arXiv:2308.07702 [cs].

[29] Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhattacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, Kai Shu, Lu Cheng, and Huan Liu. 2025. From Generation to Judgment: Opportunities and Challenges of LLM-as-a-judge. doi:10.48550/arXiv.2411.16594 arXiv:2411.16594 [cs].

[30] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*. Association for Computational Linguistics, Barcelona, Spain, 74–81. https://aclanthology.org/W04-1013/

[31] Kristen A. Lindquist and Lisa Feldman Barrett. 2008. Constructing Emotion: The Experience of Fear as a Conceptual Act. *Psychological Science* 19, 9 (Sept. 2008), 898–903. doi:10.1111/j.1467-9280.2008.02174.x Publisher: SAGE Publications Inc.

[32] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word–Emotion Association Lexicon. *Computational Intelligence* 29, 3 (2013), 436–465. doi:10.1111/j.1467-8640.2012.00460.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8640.2012.00460.x.

[33] A. Moreo, M. Romero, J. L. Castro, and J. M. Zurita. 2012. Lexicon-based Comments-oriented News Sentiment Analyzer system. *Expert Systems with Applications* 39, 10 (Aug. 2012), 9166–9180. doi:10.1016/j.eswa.2012.02.057

[34] Irina Nalis and Julia Neidhardt. 2023. Not Facial Expression, nor Fingerprint – Acknowledging Complexity and Context in Emotion Research for Human-Centered Personalization and Adaptation. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 325–330. doi:10.1145/3563359.3596990

[35] Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. 2017. Why We Need New Evaluation Metrics for NLG. doi:10.18653/v1/D17-1237 arXiv:1707.06875 [cs].

[36] OpenAi. 2024. OpenAI o1 System Card. https://openai.com/index/openai-o1-system-card/

[37] OpenAi. 2024. Structured Output. https://platform.openai.com/docs/guides/structured-outputs/how-to-use?context=with_parse#how-to-use

[38] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, and Rada Mihalcea. 2023. Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research. *IEEE Transactions on Affective Computing* 14, 1 (Jan. 2023), 108–132. doi:10.1109/TAFFC.2020.3038167 Conference Name: IEEE Transactions on Affective Computing.

[39] James A. Russell. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110, 1 (2003), 145–172. doi:10.1037/0033-295X.110.1.145 Place: US Publisher: American Psychological Association.

[40] Carlo Strapparava and Rada Mihalcea. 2007. SemEval-2007 Task 14: Affective Text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Eneko Agirre, Lluís Màrquez, and Richard Wicentowski (Eds.). Association for Computational Linguistics, Prague, Czech Republic, 70–74. https://aclanthology.org/S07-1013/

[41] Chris van der Lee, Albert Gatt, Emiel van Miltenburg, Sander Wubben, and Emiel Krahmer. 2019. Best practices for the human evaluation of automatically generated text. In *Proceedings of the 12th International Conference on Natural Language Generation*, Kees van Deemter, Chenghua Lin, and Hiroya Takamura (Eds.). Association for Computational Linguistics, Tokyo, Japan, 355–368. doi:10.18653/v1/W19-8643

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. doi:10.48550/arXiv.1706.03762 arXiv:1706.03762 [cs].

[43] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems* 35 (Dec. 2022), 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

[44] Hannes Werthner. 2024. Digital Transformation, Digital Humanism: What Needs to Be Done. *Hannes Werthner· Carlo Ghezzi· Jeff Kramer· Julian Nida-Rümelin· Bashar Nuseibeh· Erich Prem·* (2024), 115.

[45] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. 2017. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Comput. Surv.* 50, 2 (May 2017), 25:1–25:33. doi:10.1145/3057270

[46] Yifan Zhang, Yang Yuan, and Andrew Chi-Chih Yao. 2024. Meta Prompting for AI Systems. doi:10.48550/arXiv.2311.11482 arXiv:2311.11482 [cs].

[47] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. doi:10.48550/arXiv.2306.05685 arXiv:2306.05685 [cs].

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009