Automated Discourse Analysis for Attitudinal Profiling in Textual Data

Alberto Sulaiman Sade Junior,
Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil,
alberto.junior@ime.eb.br,
https://orcid.org/0009-0007-3510-9531
Cláudia Rödel Bosaipo Sales da Silva,
Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil,
rodel.claudia@ime.eb.br,
https://orcid.org/0000-0002-6694-6750
Ronaldo Ribeiro Goldschmidt,
Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil,
ronaldo.rgold@ime.eb.br,
https://orcid.org/0000-0003-1688-0586

Abstract: *Competency-based learning is a transformative approach that seeks to integrate knowledge, skills, and attitudes (KSA) (Zabala & Arnau, 2015) in the development of learners. In this context, attitudes—understood as observable behavioral tendencies shaped by affective, cognitive, and conative components—play a crucial role in shaping professional identity and decision-making. In military education, the Brazilian Army's NDACA framework formalizes strategies for attitudinal development and evaluation through structured pedagogical practices. Despite the growing interest in using Natural Language Processing (NLP) to identify behavioral traits in text, few studies focus on attitudinal profiling through discourse analysis. To address this gap, we developed a model that leverages Large Language Models (LLMs) to infer and classify attitudinal content from open-ended textual responses. Applied to responses from 14 military students enrolled in a "Leadership and Management" course, the model demonstrated promising results in detecting patterns aligned with the NDACA framework. These findings suggest that LLM-based methods may support attitudinal assessment in educational contexts (Henklein & Carmo, 2013), offering scalable and cost-effective insights into learners' values, dispositions, and behavioral trends.*

Keywords: *Automated Discourse Analysis, Attitudinal Profiling, Competency-Based Education, Large Language Models, Military Education*

Análise Automatizada de Discurso Para Perfil Atitudinal em Dados Textuais

Resumo: *A aprendizagem baseada em competências é uma abordagem transformadora que busca integrar conhecimentos, habilidades e atitudes (CHA) no desenvolvimento dos alunos (Zabala & Arnau, 2015). Nesse contexto, as atitudes — compreendidas como tendências comportamentais observáveis moldadas por componentes afetivos, cognitivos e conativos — desempenham um papel crucial na formação da identidade profissional e na tomada de decisões. Na educação militar, o referencial NDACA do Exército Brasileiro formaliza estratégias para o desenvolvimento e a avaliação de atitudes por meio de práticas pedagógicas estruturadas. Apesar do crescente interesse no uso do Processamento de Linguagem Natural (PLN) para identificar traços comportamentais em textos, poucos estudos se concentram na análise de atitudes por meio da análise discursiva. Para preencher essa lacuna, desenvolvemos um modelo que utiliza Modelos de Linguagem de Larga Escala (LLMs) para inferir e classificar conteúdos atitudinais a partir de respostas textuais abertas. Aplicado a respostas de 14 alunos militares matriculados em um curso de "Liderança e Gestão", o modelo demonstrou resultados promissores na detecção de padrões alinhados ao referencial NDACA. Esses achados sugerem que métodos baseados em LLMs podem apoiar a avaliação de atitudes em contextos educacionais (Henklein & Carmo, 2013), oferecendo percepções escaláveis e de baixo custo sobre valores, disposições e tendências comportamentais dos alunos.*

Palavras-chave: *Análise Automatizada do Discurso, Perfil Atitudinal, Educação por Competências, Modelos de Linguagem de Larga Escala, Educação Militar*

1. Introduction

Competency-based education (CBE) aims to develop learners holistically by integrating three core dimensions: Knowledge, Skills, and Attitudes (KSA). While knowledge and technical skills are often emphasized, the attitudinal dimension is essential in shaping learners' disposition to act, reflect, and engage ethically and socially. In the Brazilian military education system, attitudinal development is regulated by the *Normas para Desenvolvimento e Avaliação dos Conteúdos Atitudinais* (NDACA), which formalize pedagogical practices and behavioral indicators for fostering and evaluating attitudes in military students (Brazilian Army, 2019). These attitudes include responsibility, cooperation, discipline, and emotional balance, and are assessed through observation-based scales that reflect how students respond to training scenarios and interpersonal dynamics.

Text-based attitudinal profiling offers a promising path to enhance this evaluation process. Attitudes are often reflected in how individuals write about experiences, choices, and values. Discourse analysis enables educators and researchers to detect these signals and align them with normative expectations (Imamović et al., 2024; Yu et al., 2024). While previous work has demonstrated the feasibility of extracting emotional and cognitive traits from text (Gilardi et al., 2023), studies focusing specifically on attitudinal constructs grounded in pedagogical frameworks like NDACA remain limited (Brazilian Army, 2019).

To contribute to this emerging field, the present study explores the use of artificial intelligence—particularly Large Language Models (LLMs)—to automatically infer attitudes from open-ended textual responses. Our model analyzes written responses from 14 military students in a "Leadership and Management" course, aiming to infer attitudes based on textual indicators. These inferences are compared to scores assigned by trained human raters based on behavioral observation.

Preliminary results show strong consistency between human evaluations and more conservative patterns in LLM scoring, with moderate alignment observed in behaviorally grounded constructs such as Responsibility.

This paper is structured as follows: Section 2 presents the theoretical and regulatory background for attitudinal development. Section 3 reviews related work on NLP-based attitudinal and discourse annotation. Section 4 introduces the proposed LLM-based method. Section 5 presents the experimental results. Finally, Section 6 offers conclusions and outlines directions for future research.

2. Background

The integration of attitudinal learning within military education in Brazil is formally guided by the *Normas para Desenvolvimento e Avaliação dos Conteúdos Atitudinais* (NDACA), a regulatory framework established by the Brazilian Army's Department of Education and Culture (DECEx, 2019). According to NDACA, attitudes are relatively stable tendencies to act in specific ways toward norms or values, and they comprise three interrelated components: affective (how one feels), cognitive (what one believes), and behavioral or conative (how one acts or the willingness to act). These components shape observable behaviors and contribute to

the development of the ethical, emotional, and social competencies expected from military personnel.

NDACA outlines pedagogical strategies for fostering attitudinal development through structured interactions between instructors and students, emphasizing dialogic engagement, ethical modeling, and task-based observation. It also prescribes a detailed evaluation system based on multidimensional observation scales—conducted by instructors (vertical evaluation), peers (lateral evaluation), and self-assessment—which allows for comprehensive tracking of attitudinal progression. Each attitude is represented through descriptive behavioral indicators, known as *pautas*, which are used in exercises such as problem-solving, group work, and simulations.

This framework reflects a broader educational philosophy in which attitudinal development is inseparable from military identity and mission-readiness. As such, NDACA provides both a conceptual basis and an operational model for assessing affective and value-laden dimensions of behavior. In the context of this study, NDACA serves as the normative foundation for designing and evaluating the experimental use of LLMs to assist in attitudinal assessment—an approach that bridges military pedagogy and cutting-edge NLP technologies.

3. Related Work

The three studies highlighted here represent distinct methodological angles: attitude annotation based on Appraisal Theory (Imamović et al., 2024), comparative performance across annotation domains (Gilardi et al., 2023), and corpus-based pragma-discursive analysis using prompt engineering strategies (Yu et al., 2024). Together, these works contextualize and inform the methodological approach adopted in the present study.

Recent research has explored the use of large language models (LLMs) to support or automate linguistic annotation tasks (Landim et al., 2023), particularly in the domains of sentiment analysis, stance detection, and discourse analysis. Imamović et al. (2024) evaluated the ability of ChatGPT to perform attitude annotation based on the Appraisal Theory, using English TED Talk transcripts. Their study focused on the Attitude dimension—comprising Affect, Judgement, and Appreciation—and revealed that while ChatGPT demonstrated high precision in detecting evaluative items (94.49%), it suffered from low recall (26.74%) and struggled with fine-grained classification. Additionally, the model showed inconsistencies across runs and occasionally hallucinated evaluative expressions do not present in the original texts. These results point to both the promise and the limitations of LLMs in high-level pragmatic annotation, particularly under zero-shot prompting conditions.

Gilardi, Alizadeh, and Kubli (2023) conducted a broader evaluation of ChatGPT's annotation capabilities across multiple tasks, including relevance, stance, topic classification, and frame detection. Drawing on a corpus of over 6,000 tweets and news articles, they compared the performance of ChatGPT with that of crowd workers from Amazon Mechanical Turk. Their findings show that ChatGPT outperformed crowd workers in both zero-shot accuracy—by an average of 25 percentage points—and intercoder agreement, even exceeding that of trained human annotators. Moreover, the cost per annotation using ChatGPT was thirty times lower than MTurk, highlighting LLMs' efficiency and scalability advantages. This work underscores the potential of LLMs to transform text-annotation practices in computational social science and political communication.

In the domain of corpus linguistics, Yu et al. (2024) investigated the use of GPT-3.5 and GPT-4 for automating pragma-discursive annotation, specifically focusing on the speech act of apology. Adopting a local grammar approach, they compared LLM outputs with human annotations, tagging components such as APOLOGISING, REASON, APOLOGISER, APOLOGISEE, and INTENSIFIER. Their results showed that GPT-4 (via the Bing chatbot) reached an instance-level accuracy of 92.7%, closely approaching that of a human annotator (95.4%). Notably, GPT-4 even outperformed the human coder in identifying certain open-ended categories like REASON. The study also emphasized the importance of prompt engineering, providing a refined zero-shot prompting strategy that enhanced performance. Their work illustrates that LLMs are viable tools for corpus-based pragma-discursive analysis, offering significant time savings and scalability with only minimal human oversight.

Together, these studies illustrate the emerging capacity of LLMs to handle complex linguistic annotation tasks across domains—from applied NLP to corpus-based pragmatics. They also highlight key challenges, including prompt sensitivity, annotation consistency, and the contextual nuance required for pragmatic categories. These insights directly inform the present work, which seeks to leverage LLMs in a pedagogical context, extending current research into new methodological applications. Despite these contributions, no previous study has explored attitudinal profiling from textual responses produced in a military educational setting in Brazil.

4. Proposed Solution

This section describes the methodological solution developed to evaluate attitudinal dimensions in open-ended textual responses using Large Language Models (LLMs). The approach combines observational assessments performed by human raters with automated textual inference, organized through a five-phase workflow. Each phase contributed to a specific layer of the evaluation pipeline, from data preparation to output structuring, ensuring procedural rigor, reproducibility, and alignment with institutional norms.

4.1 Workflow Overview and Execution

The method followed a five-phase workflow, with phases developed independently but integrated sequentially for analysis:

Phase 1 – Corpus Structuring and Preparation. Raw data consisted of open-ended responses written by 14 military students in a leadership course, each responding to eight questions. These responses, originally stored in spreadsheet format, were reformatted into a document organized by `userid`, `coluna`, and `response_text`, enabling reliable parsing for the LLM. All student identities were anonymized in accordance with the Brazilian General Data Protection Law (Lei Geral de Proteção de Dados – LGPD).

Phase 2 – Human Evaluation Mapping. In parallel, two experienced raters independently assessed student attitudes based on longitudinal observation and interaction during the course. Each rater assigned scores from 1.0 to 10.0 for four constructs: Communication, Decision, Responsibility (attitudes), and Moral Courage (value). Due to the high quality of student performance, scores were concentrated between 8.0 and 10.0. The evaluations were recorded in separate spreadsheets for later comparison.

Phase 3 – Prompt Engineering and LLM Configuration. Two zero-shot prompts were designed for use with GPT-4: one targeting the NDACA-defined value Moral Courage, and another focusing on the three attitudes—Communication, Decision, and Responsibility. Both prompts were grounded in official NDACA descriptions.

Each prompt was structured to elicit a single score per construct, per student, and instructed the model to return results in tabular format sorted by `userid`. The output was formatted consistently to facilitate later comparison across evaluators.

Crucially, the evaluation strategy preserved full methodological independence between human and machine assessments. Human raters based their judgments on in-person observation of student behavior over the course of a leadership program. The LLM, in contrast, operated exclusively on the open-ended textual responses and received no training data, annotated examples, or exposure to human evaluations—ensuring a strict zero-shot configuration.

Phase 4 – LLM Execution, Calibration, and Conceptual Mapping. The prompts were executed using GPT-4, generating one score per student for each of the four target constructs: Communication, Decision, Responsibility, and Moral Courage. Initial results revealed broader dispersion than those observed in the human evaluations, prompting a recalibration process. To ensure comparability, the prompts were adjusted to constrain model outputs within a more focused interval—from 8.0 to 10.0—mirroring the empirical range used by the human raters. This recalibration improved score consistency and alignment with the observed distribution patterns.

Following score generation, the results were normalized and mapped into three conceptual bands to improve interpretability and support categorical comparison across evaluators. Although the scoring scale nominally ranged from 1.0 to 10.0, all observed results from both human and LLM evaluations clustered within the 8.0–10.0 interval. This justified the adoption of a simplified three-level classification: A (9.3–10.0): Excellent attitudinal alignment / B (8.6–9.2): Good but not outstanding performance / C (8.0–8.5): Satisfactory but limited expression.

These thresholds were empirically defined based on the actual distribution of scores and were designed to reflect NDACA's formative and non-binary approach to attitudinal evaluation. The A–B–C mapping provided a clear interpretive lens to examine agreement across evaluators, offering both numerical sensitivity and conceptual transparency.

Phase 5 – Output Structuring. The final phase of the workflow focused on consolidating the evaluation data into structured formats to support the comparative analysis presented in the next section. Two summary tables were produced to reflect the dual nature of the evaluation outputs: Table 1 (*Raw numerical scores assigned by Human 1, Human 2, and the LLM*), which displays the raw numerical scores assigned by Human Rater 1 (H1), Human Rater 2 (H2), and the LLM for each of the four constructs—Communication (C), Decision (D), Responsibility (R), and Moral Courage (MC); and Table 2 (*Data discretized into A–B–C categories*), which presents the same data in a discretized format, using the A–B–C conceptual framework introduced earlier.

In both tables, column headers follow the format X_Y, where X denotes the construct and Y identifies the evaluator. This naming convention promotes traceability and consistency in the interpretation of results across dimensions and raters.

The conceptual categories applied in Table 2—A (9.3–10.0), B (8.6–9.2), and C (8.0–8.5)—were defined empirically based on the actual distribution of scores and aligned with the NDACA's formative, non-binary approach to attitudinal evaluation. These tables offer a unified view of the numerical and categorical outputs produced by each evaluator, serving as the foundation for the analytical discussion developed in Section 5.

Table 1: Raw numerical scores assigned by Human 1, Human 2, and the LLM

#	C_H1	C_H2	C_LLM	D_H1	D_H2	D_LLM	R_H1	R_H2	R_LLM	MC_H1	MC_H2	MC_LLM
1	10,00	9,00	8,70	9,67	9,67	8,40	10,00	10,00	9,20	9,67	9,67	8,70
2	10,00	9,33	9,90	9,80	9,80	8,40	10,00	9,00	8,10	9,67	9,67	9,90
3	9,33	10,00	9,50	9,50	9,50	8,60	9,33	10,00	9,20	9,67	9,67	9,50
4	10,00	10,00	9,20	9,67	9,67	9,00	9,33	8,67	8,30	10,00	10,00	9,20
5	9,67	10,00	8,30	9,33	9,33	8,90	9,33	10,00	8,10	10,00	10,00	8,30
6	9,00	10,00	8,30	9,50	9,50	8,60	10,00	10,00	9,90	9,50	9,50	8,30
7	10,00	9,33	8,10	9,67	9,67	9,20	10,00	9,33	9,90	9,67	9,67	8,10
8	9,67	9,67	9,70	9,67	9,67	8,30	9,33	10,00	9,60	10,00	10,00	9,70
9	9,00	8,67	9,20	8,50	8,50	8,60	10,00	10,00	8,60	9,50	9,50	9,20
10	10,00	9,00	9,40	9,67	9,67	8,70	10,00	10,00	8,20	9,67	9,67	9,40
11	9,33	10,00	8,00	9,67	9,67	8,90	9,33	10,00	9,40	9,67	9,67	8,00
12	10,00	9,33	9,90	9,50	9,50	9,60	10,00	9,67	8,90	9,67	9,67	9,90
13	9,67	10,00	9,70	9,50	9,50	8,40	9,33	10,00	8,20	10,00	10,00	9,70
14	10,00	10,00	8,40	9,67	9,67	9,00	9,67	9,00	9,00	9,67	9,67	8,40

Table 2: Data discretized into A–B–C categories

#	C_H1	C_H2	C_LLM	D_H1	D_H2	D_LLM	R_H1	R_H2	R_LLM	MC_H1	MC_H2	MC_LLM
1	A	B	B	A	A	C	A	A	B	A	A	B
2	A	A	A	A	A	C	A	B	C	A	A	A
3	A	A	A	A	A	B	A	A	B	A	A	A
4	A	A	B	A	A	B	A	B	C	A	A	B
5	A	A	C	A	A	B	A	A	C	A	A	C
6	B	A	C	A	A	B	A	A	A	A	A	C
7	A	A	C	A	A	B	A	A	A	A	A	C
8	A	A	A	A	A	C	A	A	A	A	A	A
9	B	B	B	C	C	B	A	A	B	A	A	B
10	A	B	A	A	A	B	A	A	C	A	A	A
11	A	A	C	A	A	B	A	A	A	A	A	C
12	A	A	A	A	A	A	A	A	B	A	A	A
13	A	A	A	A	A	C	A	A	C	A	A	A
14	A	A	C	A	A	B	A	B	B	A	A	C

5. Result Analysis

This section is organized into two subsections. Subsection 5.1 compares the evaluators' scores numerically, and subsection 5.2 analyzes agreement based on the discretized conceptual categories.

5.1. Score-Level Comparison

This subsection analyzes the numerical scores assigned by Human Rater 1, Human Rater 2, and the LLM across all constructs. While the human scores were derived from direct behavioral observation over the course of the program, the LLM assessments were generated using zero-shot prompting applied to students' written responses.

Figures 1 through 4 display score trends for each construct. The x-axis represents student IDs (1–14), and the y-axis shows scores ranging from 8.0 to 10.0. Human 1 is shown in blue, Human 2 in green, and the LLM in red.

Figure 1 (Communication) reveals close alignment between the two human raters, with high and stable scores. The LLM assigns lower scores overall, especially for students 1, 5, 6, 7, and 11, demonstrating a more conservative scoring pattern.

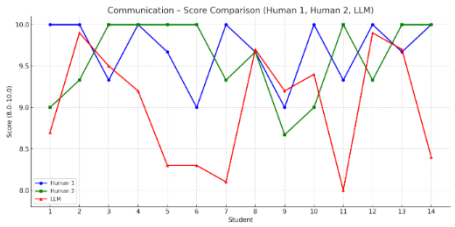


Figure 2 (Decision) shows complete agreement between Human 1 and Human 2, with identical scores for all students. The LLM again scores consistently lower and never exceeds 9.5, reinforcing its conservative bias.

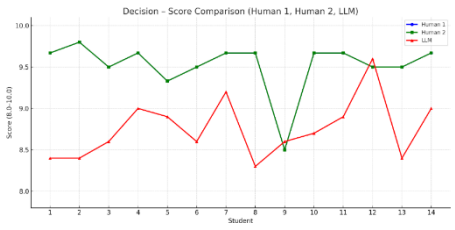


Figure 3 (Responsibility) illustrates strong alignment between the human raters, while the LLM diverges in several cases, notably for students 2, 4, 5, and 13, where it assigns lower scores.

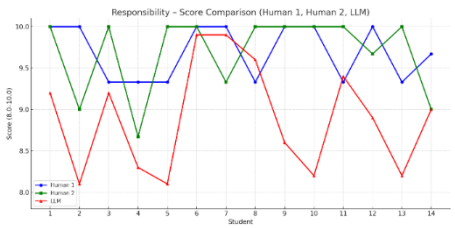
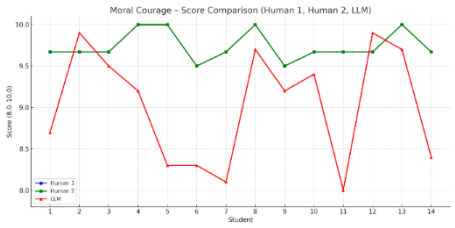


Figure 4 (Moral Courage) confirms high agreement between humans and highlights the LLM's tendency to score lower, particularly for students 1, 5, 7, and 11.



To quantify these relationships, we calculated Pearson correlation coefficients among the three evaluators. Table 3 presents these coefficients by dimension. As expected, the strongest correlations occur between Human 1 and Human 2. In contrast, correlations between the LLM and the human raters are generally low, suggesting limited linear agreement.

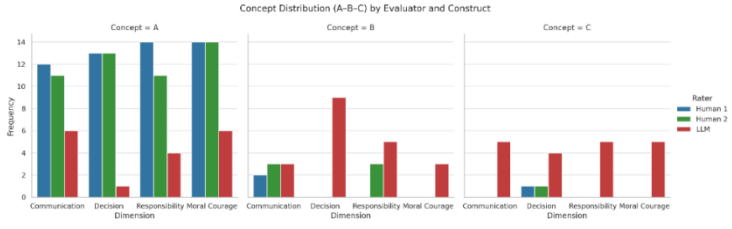
Table 3: Pearson Correlation Coefficients by Construct

Dimension	H1 vs H2	H1 vs LLM	H2 vs LLM
Communication	-0,134	0,202	-0,251
Decision	1	0,04	0,04
Responsibility	-0,06	0,129	0,202
Moral Courage	1	0,22	0,22

These results highlight a fundamental methodological difference between observation-based and text-based assessment. While human raters show high consistency—particularly in constructs with clear behavioral expression—the LLM demonstrates greater variability and a tendency toward mid-range scores.

5.2. Concept-Level Comparison (A–B–C)

To facilitate categorical comparison, the original scores were transformed into conceptual levels using the A–B–C scheme described in Section 4.3: Figure 5 presents the distribution of conceptual labels across the four constructs. Human raters show strong convergence in categories A and B, particularly in Decision and Moral Courage. The LLM, in contrast, favors categories B and C, reinforcing its conservative scoring behavior.



To assess categorical agreement, we calculated the percentage of students assigned the same concept by each pair of raters (Table 4). Full agreement between the human raters was observed in Decision and Moral Courage. Agreement between the LLM and either human rater was consistently lower, particularly in Communication and Responsibility.

Table 4: Table 4: Inter-rater agreement (%) by conceptual category

Dimension	H1 = H2	H1 = LLM	H2 = LLM	% H1 = H2	% H1 = LLM	% H2 = LLM
Communication	11	7	7	78,6	50	50
Decision	14	1	1	100	7,1	7,1
Responsibility	11	4	5	78,6	28,6	35,7
Moral Courage	14	6	6	100	42,9	42,9

These results indicate that while the LLM can approximate human categorizations in certain contexts—particularly in constructs such as Responsibility that are more behaviorally grounded—it exhibits noticeable limitations in domains requiring affective sensitivity or moral reasoning. The A–B–C classification framework serves as an effective interpretive lens for highlighting these discrepancies and identifying areas of low agreement between evaluators.

6. Conclusions

This study investigated the potential and limitations of using Large Language Models (LLMs) for attitudinal profiling within the context of Brazilian military education. All stages of the experiment—including data collection, prompting, and analysis—were conducted in Brazilian Portuguese, and the NDACA framework served as both the normative and operational basis for the evaluation process.

The proposed solution compared scores attributed by two experienced human raters—based on sustained in-person observation throughout a leadership course—with those generated by a zero-shot LLM (GPT-4), applied to students' textual responses to open-ended questions. The evaluation focused on three attitudes (Communication, Decision, and Responsibility) and one value (Moral Courage), using a five-phase workflow designed to ensure methodological independence and transparency.

Results showed high consistency between human raters, particularly in constructs grounded in behavioral observation, such as Decision and Moral Courage. The LLM produced more conservative and less varied scores overall, showing lower correlation with human evaluators—especially in constructs requiring contextual, affective, or moral interpretation. Nonetheless, it demonstrated higher alignment in the construct of Responsibility, suggesting a greater capacity for identifying clear, behaviorally anchored discourse.

These findings reinforce the notion that LLMs may serve as valuable complementary tools for formative evaluation, particularly in scenarios where human evaluators are unavailable or scarce. When embedded within structured educational frameworks such as NDACA, LLMs have the potential to support attitudinal assessment processes that are efficient, scalable, and ethically grounded.

Building on the results presented in this study, future work should pursue refinements in prompt construction, model configuration, and human-machine integration strategies. Extensions include:

- Few-shot or rubric-based prompting: Incorporating annotated examples or task-specific rubrics may improve the model's sensitivity to affective, moral, and context-dependent dimensions.

- Multimodal assessment models: Future studies could combine text-based inference with structured behavioral observations to generate more comprehensive attitudinal profiles.

- Fine-tuning with institutional corpora: Adapting LLMs using annotated educational materials from military training environments may improve domain specificity and consistency with institutional expectations.

- Responsible integration frameworks: Research should explore how AI-based assessments can be embedded into pedagogical practice in transparent, fair, and norm-compliant ways—especially when used to complement human evaluations rather than replace them.

Taken together, these directions aim to bridge the gap between the scalability of LLM-based solutions and the human depth of behavioral insight, advancing the responsible and effective use of AI in military and educational contexts.

Acknowledgments

We thank Sarah Vitória Luiz Vanderei for her assistance in the final formatting of the article.

References

BRAZILIAN ARMY. (2019). *Normas para desenvolvimento e avaliação dos conteúdos atitudinais (NDACA – EB60-N-05.013)* (3rd ed.). Rio de Janeiro: Departamento de Educação e Cultura do Exército.

GILARDI, F., ALIZADEH, M., & KUBLI, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences of the United States of America, 120*(30), e2305016120. https://doi.org/10.1073/pnas.2305016120

HENKLEIN, M. H. O., & CARMO, J. S. (2013). Contributions of behavior analysis to education: An invitation to dialogue. *Revista Brasileira de Análise do Comportamento, 9*(2).

IMAMOVIÉ, M., DEILEN, S., GLYNN, D., & LAPSHINOVA-KOLTUNSKI, E. (2024). Using ChatGPT for annotation of attitude within the Appraisal Theory: Lessons learned. In *Proceedings of the 18th Linguistic Annotation Workshop (LAW-XVIII)* (pp. 112–123). Association for Computational Linguistics. https://aclanthology.org/2024.law-1.11

LANDIM, G. P. P., TRESSO, G. J., & PASSERINI, J. A. R. (2023). Sentiment identification in texts using the TF-IDF model. In *Proceedings of the 7th Technology Symposium of Fatec Jales*. ISSN 2595-2323.

SILVA, C. S., MOREIRA, T. O., FERNANDES, I., PASSOS, C., DUARTE, J. C., & GOLDSCHMIDT, R. R. (2023). Intelligent tutoring systems in competency-based learning: A systematic literature review. *Unpublished manuscript*.

YU, D., LI, L., SU, H., & FUOLI, M. (2024). Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis: The case of apologies. *International Journal of Corpus Linguistics*. Advance online publication. https://doi.org/10.13140/RG.2.2.26420.48001

ZABALA, A., & ARNAU, L. (2015). *How to learn and teach competencies*. Penso Publishing.