

Pragmatic users and skeptical nonusers: A qualitative typology of ChatGPT adoption in physics education

Eleonore Becker¹, Antonia Bauer², Johanna Schrader³,
Philipp Bitzenbauer⁴, and Joaquin Veith⁴

¹*Leipzig University, Institute of Physics Education, Vor dem Hospitalore 1, 04103, Leipzig, Germany*

²*Friedrich-Alexander-University Erlangen-Nürnberg,*

Physics Education Research Group, Staudtstraße 7, 91058, Erlangen, Germany

³*Leibniz University Hannover, L3S Research Center, Hannover, Germany*

*and CAIMed—Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine
Hannover, Germany*

⁴*Leipzig University, Institute of Physics Education, Vor dem Hospitalore 1,
04103, Leipzig, Germany*



(Received 14 November 2025; accepted 4 June 2026; published 6 August 2026)

Physics students' use of ChatGPT or other generative AI tools is a topic broadly discussed in physics faculties around the globe, with discussion often outpacing evidence. What remains unclear is how physics students truly engage with these tools and by what perspectives their choices are shaped. This study develops a data-driven typology of how students perceive the tool's usefulness and its risks. Through a qualitative content analysis of 1189 survey responses, followed by latent class analysis (LCA), we identified two distinct user profiles. The majority (70%) are "Pragmatic Users" who, while aware of the tool's inaccuracies, use it for scaffolding tasks like conceptual clarification and finding starting points for problem-set questions. The minority (30%) are "Skeptical Non-Users," who avoid the tool over concerns about overreliance and its potential to hinder independent problem-solving. These profiles show students making a calculated trade-off between usefulness and risk, suggesting one-size-fits-all policies on AI are inadequate. Thus, rather than promoting uniform adoption, a differentiated pedagogy could create space for students to make informed, self-determined choices—supporting reflective AI use for those who opt in, while equally validating and enabling the learning practices of those who choose not to use generative AI.

DOI: [10.1103/gdlw-1637](https://doi.org/10.1103/gdlw-1637)

I. INTRODUCTION

The rise of generative artificial intelligence (AI) has been most profound among young adults, with the 18–24 age cohort consistently emerging as its most active user base [1–4]. This demographic trend has direct and significant implications for higher education. Recent surveys confirm widespread adoption in academic settings, with reports indicating that between 61% and 88% of university students now use ChatGPT for their studies [5–10].

Across all studies, ChatGPT has the highest usage rates [1,3,4,11], consistently exceeding its nearest competitors by over 20 percentage points [1,3,10]. This popularity is mirrored in academia, where it is the most frequently used tool by students [5,6,10,12] and consequently the primary focus of most academic research on this topic [13–15].

This research focus highlights a key methodological consideration. Generative AI models are not interchangeable; built on distinct architectures and training data, they possess unique capabilities and limitations, with strengths including conversational dialogue (ChatGPT), cross-model reasoning (Gemini), or code generation (Copilot) [16]. Consequently, findings from one tool may not be generalizable to generative AI as a whole, and a tool-specific approach can yield more precise and interpretable results.

However, the existing body of ChatGPT research, while tool specific, remains largely interdisciplinary [17–19] or focused on fields such as social sciences, humanities, and medical education [20–22]. This imbalance leaves a critical research gap in science, technology, engineering, and mathematics (STEM) disciplines such as physics, where learning depends on problem-oriented and conceptual thinking [23–25]. A tension exists between these learning requirements and ChatGPT's documented performance: while the tool performs well on solving well-defined, text-based problems [26–29], it falters on tasks requiring physical modeling, multistep reasoning, and

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International* license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

spatial conceptualization [27–30]. Despite this apparent misalignment, adoption among physics students has surged from minimal usage [31] to approximately 75% [32]. This situation raises the question: How and why are physics students integrating a tool with known limitations into their individualized learning processes? To investigate this, we turn to established theories of technology adoption. Models such as the technology acceptance model (TAM) [33] and the unified theory of acceptance and use of technology (UTAUT) [34], along with their extensions, are widely used to explain technology adoption in educational settings [cf. [35–39]]. Drawing on these frameworks, this research identifies distinct user types to better understand how physics students reconcile the tool’s perceived benefits with its known limitations for discipline-specific learning.

II. RESEARCH BACKGROUND

The following section provides the theoretical and empirical context for this investigation across three levels of granularity. Section II A reviews foundational frameworks of technology acceptance, specifically the TAM and the UTAUT. Section II B synthesizes empirical findings on ChatGPT adoption in higher education, focusing on primary predictors of usage, the role of perceived risk, and the emergence of person-centric user typologies. Finally, Sec. II C examines the discipline-specific landscape of ChatGPT usage in physics education, highlighting the pedagogical tensions and student engagement strategies relevant to the field.

A. Technology acceptance

Research on technology acceptance seeks to explain why individuals adopt or reject technological innovations. Two of the most prominent frameworks, the TAM [33,40–42] and the UTAUT [34], share the underlying principle that multiple factors jointly influence the single construct behavioral intention, which in turn is the strongest predictor of actual user behavior [33,34,41,43,44].

TAM posits that perceived usefulness (PU) and perceived ease of use (PEOU) shape users’ behavioral intentions [33,40–42]. PU refers to “the degree to which a person believes that using a particular system would enhance his or her job performance,” whereas PEOU describes “the degree to which a person believes that using a particular system would be free of effort” [33] (both p. 320). UTAUT builds upon these by synthesizing a total of eight models into a single framework, mapping TAM’s constructs to performance expectancy (PU) and effort expectancy (PEOU) [34]. UTAUT also introduces two additional determinants: social influence, the perceived pressure from important others to use a tool, and facilitating conditions, the perceived availability of technical and organizational support [34]. These additional

constructs provide a theoretical basis for the research on students’ adoption of ChatGPT discussed in Sec. II B.

While TAM and UTAUT have proven robust across diverse settings, they have frequently been extended and modified numerous times [42,45–50]. To tailor these frameworks to the specific case of AI, researchers have introduced new constructs [cf. [51–56]]. For instance, models have been expanded to include technical attributes like transparency [52] and explainability [52], as well as psychological and ethical factors such as trust [51,53,55,57,58], moral obligation [57], and (perceived) anthropomorphism [52,58]. Notably, perceived risk (the potential for negative consequences) has emerged as a barrier to adoption in AI-assisted learning [51,55]. Building upon these frameworks, this research employs perceived usefulness and perceived risk to establish a qualitative typology of student adoption. A more detailed discussion of how these constructs informed our methodology is provided in Sec. IV B.

B. Acceptance and usage of ChatGPT

Research applying extended frameworks to the study of ChatGPT in higher education indicates that while classic predictors remain relevant, the nature of AI-generated content brings constructs like trust and perceived risk to the forefront. Performance expectancy and PU have consistently been found to predict behavioral intention [15,51,54,59–62]. Similarly, habit has also been identified as a significant predictor [15,54,59]. In contrast, findings are less conclusive regarding the influence of other factors, including effort expectancy and PEOU [15,51,60–62], social influence [15,51,59–62], and hedonic motivation [15,51,54]. Facilitating conditions were generally found not to influence behavioral intention [15,54,59,61,62], although one study identified a direct link to actual use behavior [61]. Finally, a positive attitude toward the technology has consistently been shown to influence behavioral intention [13,15,51,60]. The overall importance of utility is reflected in the tool’s most common applications, which include research and information gathering [5–7], clarifying conceptual questions [5,6], and generating ideas [5].

In response to students’ frequent encounters with incomplete, false, and irrelevant AI-generated content [5], technology acceptance frameworks have been adapted to account for the influence of trust and perceived risk. Generally, trust has been shown to positively influence student attitudes and behavior intentions [51,55,63], although Dietrich and Grassini [15] did not find a significant relationship. Adding complexity to this view, Shahzad *et al.* [64] found that trust mediates the relationship between a student’s self-efficacy and their subsequent usage outcomes, an effect moderated by perceived usefulness and ease of use. This finding suggests that a student’s confidence in their ability to manage the

technology's limitations is a key factor in its adoption. Multiple studies report a positive relationship between students' perceived risks and their intention to use ChatGPT [51,55]. These researchers interpret this phenomenon as a "calculated trade-off" rather than recklessness, suggesting that students who are most aware of the technology's pitfalls are precisely the ones who have consciously weighed those risks against the perceived benefits [51,55].

However, the traditional, variable-centered application of these models assumes that a single set of predictive relationships applies to all students, an approach that can mask the presence of distinct subpopulations [65]. To address this, person-centric research seeks to identify distinct user archetypes, recognizing that predictive factors vary based on traits such as "AI digital ability" [66]. At a psychological level, students can be grouped into broad mindsets, including "AI Advocates" (48%) and skeptical "Cautious Critics" (21%) [67]. These mindsets manifest in distinct behavioral strategies; for instance, some students are "Knowledge Seekers" (17%), using AI as a cognitive learning tool, while others are "Assignment Delegators" (23%), who use it primarily to complete assignments [14]. In the specific context of programming, another study found that a majority of students (59%) adopt a "Passive Solver" approach, accepting AI-generated solutions without modification [68].

C. Acceptance and usage of ChatGPT for learning physics

Differences in students' usage of ChatGPT for learning activities underscore the need for discipline-specific analysis: STEM students use AI for conceptual assistance much more frequently than their non-STEM peers (41% versus 24%), who use it primarily for writing support (25%) [69]. These disciplinary patterns extend beyond usage to perception. For instance, even within STEM, factors such as perceived social influence can vary significantly between majors such as chemistry and biology programs compared to physics and informatics programs [70].

Within STEM, the most common use cases for generative AI in learning involve explaining complex topics [12,69,71], summarizing texts [12,71], and help with answering problem-set questions [12]. Research on student interactions with ChatGPT in physics has largely focused on problem-solving, a context where differing engagement strategies significantly impact learning outcomes [23,25,72]. On one hand, students frequently rely on low-effort, less-effective strategies when solving problems with ChatGPT without guidance. Krupp *et al.* [25] found that without guidance, students frequently default to "copy & paste" tactics, leading to significantly worse performance than control groups and a high false-positive rate, where incorrect AI answers are uncritically accepted.

This reliance on direct solutions is corroborated by Wang *et al.* [12], whose small-scale survey of STEM students revealed that a slight majority (54%) preferred to use generative AI to solve problems directly rather than as a tool to support their problem-solving process. On the other hand, structured pedagogical environments encourage students to adopt inquiry-based strategies. Interventions such as forbidding direct copying or using custom-prompted chatbots successfully shift student behavior toward constructive uses, including answer verification, conceptual inquiry, and ongoing tutorial support [23,73]. Such inquiry-based approaches are linked to improved student performance, especially when students ask conceptual rather than procedural questions [23]. These findings, along with broader evidence of learning gains from AI chatbots in educational settings [74,75], suggest that the pedagogical value of ChatGPT in physics education depends on student engagement as framed by the educational implementation.

Underpinning these divergent engagement strategies is a widespread yet pragmatic student acceptance. The overall acceptance and perception of ChatGPT in STEM [70] and specifically physics [32,75] is positive, with students viewing the tool as a valuable supplement [23,31,73]. TAM studies consistently identify perceived ease of use as a primary predictor of students' behavioral intention and actual user behavior [32,70]. Perceived usefulness is also key, as students value the tool's ability to offer instant feedback [73], provide strategic guidance [23,75], promote understanding [23,73], and help in completing tasks more quickly [12,76]. However, the quality of the interaction itself appears to shape this perceived usefulness. For instance, students who develop more detailed prompting strategies report a more positive attitude [71]. Furthermore, low-frequency users have reported significantly higher satisfaction than high-frequency users, suggesting that students perceive more value in using the tool for specific, targeted tasks [75].

This pragmatic acceptance is, however, balanced by skepticism. A primary student concern is the tool's potential for inaccuracy and errors [12,31,73,76]. Furthermore, students harbor strong concerns about over-reliance and the potential for ChatGPT to undermine their critical thinking skills [73,76], a concern that mirrors those held by educators [12,24]. This culminates in the belief that while useful, the tool may not ultimately improve grades [76] or serve as a viable alternative to traditional learning methods [71].

The tension between use and skepticism is corroborated by quantitative data. For example, a recent cross-national survey of 804 physics students by Agyare *et al.* [32] utilized an extended version of the TAM to analyze how students' perceptions of ChatGPT's usefulness and risk influence user behavior and subjective learning outcomes, defined here as the students' self-perceived gains in knowledge and skill. While students' ethical and

pedagogical concerns were found to significantly reduce their user behavior, these same concerns did not negatively moderate the subjective learning outcomes for those who did use the tool. Conversely, their findings suggest that user behavior correlates strongly with subjective learning outcomes. This result contrasts with other research, suggesting that targeted, lower-frequency engagement is more satisfying and effective (cf. [75]).

III. RESEARCH RATIONALE

The current research on ChatGPT in education is fragmented, with studies often limited by their methodology, context, or scale: On one hand, large quantitative surveys investigate acceptance and use but rely on closed-ended questions that primarily yield usage metrics [15,32,51,54,59–62]. On the other hand, insightful qualitative studies often examine tool use in settings that limit student autonomy [23,73], such as interactions with a custom-designed AI tutor [73] or during a formal, graded assessment [23]. Furthermore, many studies lack the scale [12] or domain specificity [5,69,71]. As established in the introduction, this creates a gap in understanding ChatGPT usage in physics, where a clear tension exists between the discipline’s learning demands [23–25] and the tool’s documented limitations [12,28–30]. Responding to calls for large-scale validation of findings [12] and nuanced, physics-specific investigations [32], this study aims to address these shortcomings by presenting a large-scale, open-format investigation into the autonomous use of ChatGPT by physics students. Our objective is to develop a typology of how physics students integrate ChatGPT into their learning processes. Therefore, we conducted a large-scale cross-sectional survey using an open-ended questionnaire designed to operationalize key theoretical constructs from the technology acceptance literature.

Our inquiry focuses on two constructs: perceived usefulness and perceived risk. Perceived usefulness, a predictor of behavioral intention [15,51,54,59–62], refers to the degree to which students believe a tool will benefit their learning. Conversely, perceived risk acts as a barrier to adoption [51,55] and encompasses concerns about the tool’s limitations and potential negative consequences, such as generating inaccurate content or fostering overreliance [51,55,73,76]. Together, these constructs frame a utility-risk analysis that shapes student adoption. Research suggests that students navigate this tension by making a “calculated trade-off,” weighing the tool’s perceived usefulness against the perceived risk [51,55].

To provide a snapshot of contemporary student perspectives, we address two primary research questions:

RQ1: What aspects of perceived usefulness and perceived risk characterize physics students’ views on ChatGPT in their learning?

RQ2: What distinct user types can be identified based on students’ views of ChatGPT as a learning resource in physics?

IV. METHODS

A. Study design and sample

The questionnaire was administered digitally via LimeSurvey to $N = 1189$ students ($N_1 = 800$ male, $N_2 = 356$ female, $N_3 = 33$ diverse) from German universities (a detailed description is provided in Sec. IV B). All participants were enrolled in a study program with physics as a major subject (e.g., B.Sc. physics, M.Sc. physics, or physics teacher training). Data collection was conducted during the winter term 2024/2025. Participation was voluntary, not compensated, and informed consent was obtained from all participants prior to data collection.

B. Questionnaire

Our questionnaire consisted of four open-ended questions derived from the core facets of the TAM:

Q1: In which areas have you experienced inaccurate results when working with ChatGPT in physics?

Q2: To what extent does the use of ChatGPT support the learning process in physics?

Q3: To what extent does the use of ChatGPT hinder the learning process in physics?

Q4: Provide specific examples of how you use ChatGPT as a learning resource in physics.

Questions Q2 and Q4 investigate perceived usefulness. Since “usefulness” in physics is a multifaceted concept, including aspects such as problem-solving assistance and conceptual clarification [23], these questions are complementary in eliciting students’ own definitions of usefulness, probing both the general extent of support (Q2) and concrete applications (Q4).

Conversely, questions Q1 and Q3 investigate perceived risk. We unpack this construct into two distinct dimensions to add nuance to findings that show a positive relationship between students’ overall perceived risk and their intention to use ChatGPT [51,55]. The first dimension, pedagogical risk, is the concern that the tool may harm the learning process itself by fostering overreliance or hindering critical thinking [73,76]. We target this dimension by asking about hindrances to the learning process (Q3). The second dimension, system risk, relates to the tool’s inherent reliability and the concept of perceived trust. Given the well-documented tendency for AI to generate flawed physics content [12,31,73,76], we probe for this risk by asking about direct experiences with inaccuracies (Q1).

The questionnaire was set up specifically so that participants did not have to answer all questions.

C. Data analysis

Research question 1 was investigated by means of a qualitative content analysis. Given the substantial volume of open-ended responses, we employed a two-step semi-automated qualitative content analysis approach using a generative language model (GPT-4.1 [77]). This decision was primarily motivated by considerations of scalability and analytical consistency. Recent empirical studies have demonstrated that large language models (LLMs) can achieve substantial agreement with human coders in qualitative coding tasks, particularly when prompts are carefully structured and coding procedures are clearly specified [78,79]. Our approach for qualitative content analysis was based on Merriam's principles [80] which consist of three steps: First, categories are constructed by recording emerging themes in the responses relevant to the research questions. This step is iterated until no new themes emerge throughout all responses. In a second step, the constructed codes are sorted, refined, and regrouped into broader categories. Lastly, the validity of the codes is investigated via interrater agreement. Thus, in the first step, the generative language model was prompted to generate an initial codebook (cf. Sec. IV C 1) for each survey question. The authors iteratively refined the prompt and reviewed the resulting codebooks until thematic coherence and conceptual clarity were achieved. In the second step, the finalized codebooks were used by the model to annotate the corresponding responses. To ensure methodological rigor, we complemented this automated coding with human review of a random subset of the data (cf. Sec. IV C 2).

Since the survey and responses were in German, all prompting and coding were conducted in German to avoid translation ambiguities and preserve semantic accuracy [81]. The code we used for developing the codebooks and the coding process as well as the final codebooks are available via [82].

1. Codebook development

For each survey question, we supplied the GPT-4.1 model with the question text and all associated responses as input. We prompted the model to generate a structured codebook in CSV format containing columns for a concise code label, a brief description of each code, coding rules describing when to apply the code, and a citation from the responses as an anchor example illustrating its use. The model was instructed to summarize semantically similar answers into a single code, with each code describing only one concept. Repetitions or overlapping codes were to be avoided, and the number of assigned codes was to be limited to those necessary to cover all relevant content. The codes were to be based solely on the answers that were actually present in the dataset. Finally, the model was instructed to return one code per row in a CSV-like structure, exemplified by the column names to return.

The prompt used for codebook generation was iteratively refined through discussions among the authors, who reviewed draft outputs and provided feedback until the codebooks met expectations for thematic coherence and conceptual clarity. After the automated generation phase, two of the authors manually reviewed the resulting codebooks to correct inconsistencies or merge overly specific codes, but no changes were necessary at this stage.

Because generative AI models can produce different answers to the same prompt, we used the temperature parameter T to manage the model's output stability. The temperature of the model regulates the diversity-coherence trade-off in autoregressive text generation [83]. Technically, it controls the "softness" of the probability distribution used to select the next token [84]. A low temperature ($T < 1$) favors high-probability, high-coherence tokens, resulting in a more deterministic output, while higher temperature values encourage more creativity. For the codebook development, we followed Ref. [85] and set the model's temperature parameter to $T = 0.2$. This low-temperature sampling approach was chosen over greedy decoding ($T = 0.0$), which, while deterministic, has been shown to lead to "bland and strangely repetitive" text degeneration in generative tasks [83]. Our setting of $T = 0.2$ thus prioritizes consistent and coherent summarization while still allowing for minor, high-confidence variations. All other adjustable parameters were left at their default API values [86].

We manually defined a code 0, "Not applicable," with the description "Response does not contain relevant content or does not fit into any code." This code was to be used when a response was empty, incomprehensible, or irrelevant to the question.

2. Coding process

In the second step, the finalized codebooks were used to annotate the corresponding survey responses. All empty or missing responses were automatically assigned code 0 (Not applicable). To ensure efficient and consistent annotation quality, responses were processed in batches of ten. This batch size was chosen to prevent model degradation and "lazy" annotation behavior over long sequences, as smaller batch sizes have been found to yield more stable output quality in large language models [87].

The model was prompted to assign only codes that are clearly supported by the explicit content of each response and to treat all responses independently. When multiple codes were applied, the model was instructed to assign all relevant codes but to prioritize the most specific applicable codes. Code 0 was to be assigned if the response did not meaningfully answer the question, fell outside the thematic coverage of the codebook, or consisted only of placeholder text such as "/". To further improve reliability, the prompt included several example responses paired with their correct code assignments (few-shot learning [88]), following best practices in qualitative AI annotation [89].

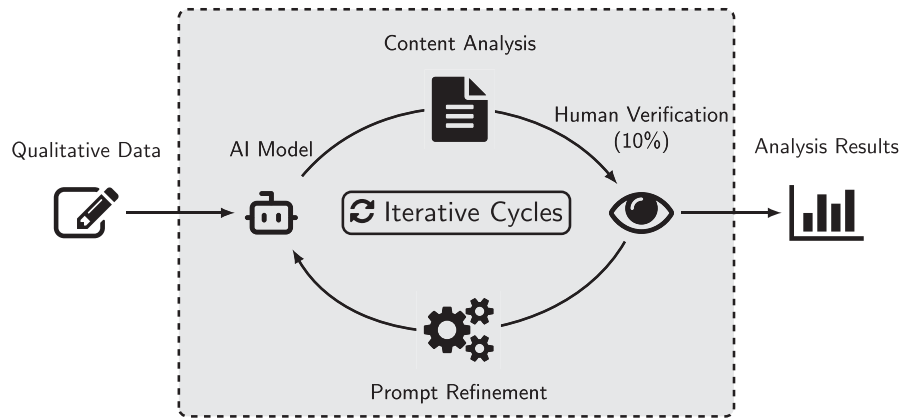


FIG. 1. Iterative workflow of the qualitative content analysis presented in this study. Qualitative questionnaire responses were processed through iterative cycles of AI-driven content analysis using GPT-4.1, human verification of a random 10% subsample, and prompt refinement to ensure accuracy and thematic coherence in codebook development and response annotation. Icons in this figure are sourced from the fontawesome and lucide-icons LaTeX packages.

As this annotation step is a classification task for the model, we favor greedy decoding. Thus, the temperature parameter for this step was set to $T = 0.0$ to ensure maximum determinism and reproducibility in the coding process [89], while all other parameters were kept at their default settings [86]. As in the previous step, the annotation prompt underwent several rounds of refinement, guided by the authors' evaluation of annotation quality: After each annotation, we randomly selected 10% of the data and checked whether each category was applied correctly. In the case of one or more erroneous annotations, the errors were analyzed and the prompt was subsequently revised to avoid those errors. These adjustments were repeated until no inaccuracies were detected—the development process is visualized in Fig. 1. Results of the annotation task were analyzed by means of frequency analyses.

3. Latent class analysis

To further analyze the qualitative data and address RQ2, we employed latent class analysis (LCA). LCA is a segmentation method that uses “maximum likelihood estimation to fit a hypothesized model to the data, based on preselected indicators, and groups participants into latent classes” [90] (p. 5). Thus, each class represents a specific response pattern, expressed via item-response probabilities [91]. In simple terms, LCA assumes that the observed response patterns arise from a small number of latent subgroups in the population. The method estimates both the size of these subgroups and the probability that members of each subgroup provide particular responses. We chose LCA over other clustering methods such as k modes due to its low sensitivity to outliers [92]. To obtain the required format of nominal data [cf. [93]], we transformed each category from the content analysis into a dichotomous variable (1 = participants' response got

assigned this category, 0 = participants' response did not get assigned this category). To ensure local independence, correlations between variables were checked prior to analysis. For model selection, we followed established guidelines by investigating several fit statistics (cf. [94,95]): First, we investigated the overall model fit using the likelihood-ratio chi-square statistic (G^2) as well as the normalized entropy (E) alongside information criteria such as the Bayesian information criterion (BIC) and the Akaike information criterion (AIC), which directly include a penalty for the number of estimated parameters to account for a balance between model fit and parsimony. Generally, an optimal class solution is evidenced by the lowest values of G^2 , AIC, and BIC [96], accompanied by values of E above 0.80, which indicate that latent classes are highly discriminating [97]. Second, we followed Ref. [90] by also investigating the consistent Akaike information criterion (CAIC) and the sample-size adjusted Bayesian information criterion (ABIC). These indices apply alternative penalty functions for model complexity, with CAIC imposing a stronger penalty than BIC and ABIC providing a milder, sample-adjusted penalty. Hence, including these additional parsimony measures offers a more nuanced assessment of model fit across varying complexity levels [91].

Finally, to improve the robustness of the LCA and reduce model complexity, only categories mentioned by at least 5% of participants were included. Participants who named only excluded categories were removed from the analysis to avoid rows consisting entirely of zeros, which could distort class estimation. This approach allows a reduction in the number of parameters to be estimated, thereby enhancing model stability and interpretability while retaining the most informative response patterns for class identification. Sample size after this preprocessing was $N = 1019$. It is important to note that this preprocessing did not affect the qualitative content

analysis prior to LCA. LCA was performed using R version 4.4.2 and the R package poLCA [98].

V. RESULTS

A. Results of the qualitative content analysis

The results of the qualitative content analysis are summarized in Tables I–IV, addressing the four questions Q1 to Q4, respectively.

1. Results regarding experienced inaccuracies with ChatGPT when learning physics

Table I presents an overview of all categories emerging in responses to question 1 (“In which areas have you experienced inaccurate results when working with

ChatGPT in physics?”), including the total numbers with which they occurred and the corresponding relative frequencies. The two most often mentioned categories refer to incorrect calculations (ID 1_1, 35%) and errors in specific or complex problems (ID 1_7, 24%), hinting at ChatGPT’s perceived limitations when it comes to computation or grasping more complex physical problems. For the former, participants reported “Horrendous calculation errors in astronomy problems” and “Linear algebra and everything related to calculations.” For the latter, responses included “Ultrafast optics, gravitational wave detection, theoretical QM (basically anything that goes beyond the double-slit experiment),” “particle physics,” and “laser physics.” In contrast to these more cognitively demanding tasks, 12% of responses also touched on experiences with

TABLE I. Categorization matrix for the first question (Q1: “In which areas have you experienced inaccurate results when working with ChatGPT in physics?”), including anchor examples as well as the total number of respondents who were assigned the respective category. The number in the last column indicates the relative frequency with which the respective category has been assigned when excluding all participants who were assigned “not applicable.” It is important to note that some responses had multiple codes applied, leading to a percentage total greater than 100%.

ID	Code <i>Example</i>	Total No.	rel # (%)
1	Incorrect calculations “ChatGPT occasionally does incorrect calculations.”	273	35
2	Incorrect or incomplete derivations “Derivations, complex mathematics.”	118	15
3	Incorrect formulas or definitions “Everything that requires formulas is often incorrect or not correctly applied.”	92	12
4	Incorrect or contradictory explanations “The answer depends on how the question is phrased. Negating a question can lead to a different answer.”	66	8
5	Incorrect programming solutions “Programming: Incorrect or circuitous solutions.”	45	6
6	Incorrect or missing sources “Publications are cited incorrectly, falsely described, and unreliable.”	21	3
7	Errors in specific or complex problems “Calculation errors occur more frequently in specific calculations.”	190	24
8	Consistency problems and varying answers “Especially when you ask the same question twice, you often get different answers.”	28	4
9	Incorrect assumptions or interpretations “Sometimes it misunderstands the task, and especially when ChatGPT really has to calculate numbers, it tends to make mistakes.”	37	5
10	Hallucinations and fabricated content “When searching for specific movie scenes, ChatGPT started inventing movie scenes that don’t exist.”	26	3
11	Superficial or simplified answers “Many topics are summarized in a simplified manner. The cutoff point of the LLM prevents newer insights if one relies solely on ChatGPT.”	29	4
12	Little or no personal experience “Never, because I rarely use ChatGPT.”	48	6

TABLE II. Categorization matrix for the second question (Q2: “To what extent does the use of ChatGPT support the learning process in physics?”), including anchor examples as well as the total number of respondents who were assigned the respective category. The number in the last column indicates the relative frequency with which the respective category has been assigned when excluding all participants who were assigned “not applicable.” It is important to note that some responses had multiple codes applied, leading to a percentage total greater than 100%.

ID	Code	tot #	rel # (%)
	Example		
1	Quick responses “It often saves me time that I would otherwise have to spend on lengthy Google searches.”	149	24
2	Individual responses “Content can be explained in such a way that it becomes understandable to specific individuals.”	62	10
3	Low threshold “You can also ask questions without embarrassment that you would not otherwise dare to ask.”	27	4
4	Promoting understanding “Often explains things better than lecturers or literature and helps when you get stuck.”	211	35
5	Problem-solving starting point “ChatGPT helps find a starting point that can be used for further calculations.”	155	25
6	Summaries and overview “ChatGPT is very good at summarizing models clearly and highlighting their advantages and disadvantages.”	71	12
7	Questions and dialogue “The opportunity to engage and ask questions at every step of the learning process is the key advantage for me.”	76	12
8	Alternative explanations “It’s good to have things explained to you again in different words.”	58	10
9	Results check “Reviewing results helps to find a starting point for tasks.”	36	6
10	Introduction to new topics “ChatGPT offers a really good start when first getting to grips with a new topic.”	15	2
11	Research support “Source search can work extremely well.”	38	6
12	Assistance with tasks “Help with solving problems. ChatGPT can help provide a starting point.”	49	8
13	Programming support “ChatGPT can be very useful, especially when it comes to coding problems.”	23	4
14	Repetition and practice “Offers the opportunity to practice individually and receive immediate feedback.”	22	4
15	Motivation and frustration relief “You don’t stagnate for so long.”	12	2
16	Promote critical thinking “You learn by evaluating the answers.”	25	4
17	Availability and flexibility “Always have access to someone who can answer questions.”	24	4
18	Language support “I mainly use it for grammar, language, rephrasing, shortening... in other words, text edits.”	22	4
19	Promoting autonomous learning “It provides a certain degree of autonomy when learning content.”	7	1

TABLE III. Categorization matrix for the third question (**Q3**: “To what extent does the use of ChatGPT hinder the learning process in physics?”), including anchor examples as well as the total number of respondents who were assigned the respective category. The number in the last column indicates the relative frequency with which the respective category has been assigned when excluding all participants who were assigned “not applicable.” It is important to note that some responses had multiple codes applied, leading to a percentage total greater than 100%.

ID	Code <i>Example</i>	tot #	rel # (%)
1	Less independence “You forget how to search for information independently and rely completely on ChatGPT.”	73	39
2	Lack of understanding of the problem “Sometimes you stop trying to understand problems because you can solve them with ChatGPT anyway.”	22	12
3	Incorrect or unreliable information “Incorrect data.”	52	28
4	Uncritical acceptance “Results are accepted uncritically and often not checked.”	26	14
5	Superficial or abridged presentation “In my opinion, the questions I can generate for my lessons using ChatGPT are not interesting.”	13	7
6	Reduced motivation or perseverance “You potentially give up too quickly on finding approaches and thus learn significantly less in the long term.”	16	8
7	Lack of research skills “People are not trained in the use of reliable sources of information, which is an important skill.”	14	7
8	Illusion of understanding “By simplifying the summary of topics, ChatGPT also creates an illusion of understanding, but it doesn’t convey the entire knowledge.”	9	5
9	Dependence on ChatGPT “If you develop a strong dependence on ChatGPT, then you have to rely on AI forever.”	6	3
10	No or little learning effect “In my opinion, ChatGPT users learn nothing except how to ask clever questions that the AI will answer.”	12	6

incorrect formulas and definitions (ID 1_3), such as “making up equations” and “using incandescent bulbs as ohmic resistors,” indicating problems with purely declarative knowledge tasks. Only a few mentions were made regarding technical aspects of LLMs, with about 3% mentioning hallucinations and fabricated content (ID 1_10), illustrated by statements such as “I was often told false statements or even completely fabricated statements” and “ChatGPT started inventing movie scenes that don’t exist to fulfill my wishes.” Another 4% of responses alluded to consistency problems and varying answers (ID 1_8), e.g., “He tells you something, you ask, “Are you sure?” He replies, “You’re right, I’m sorry, you’re right, it’s this and that” (I didn’t even tell him what I think).”

2. Results regarding the perceived usefulness of ChatGPT for physics learning

Answers to question 2 (“To what extent does the use of ChatGPT support the learning process in physics?”) revealed a comparatively broad spectrum of perceptions

(cf. Table II). Here, responses indicate that students particularly value ChatGPT for promoting understanding (ID 2_4, 35%), illustrated by examples such as “It often explains things better than the lecturer or the literature and helps when you get stuck” and “understanding concepts better.” Furthermore, participants appreciated ChatGPT providing an initial approach to problem solving (ID 2_5, 25%), thus enabling them to get started without extensive prior research, using ChatGPT to “find solutions and have complex issues explained to you,” “get inspiration for solving problems,” and “offer a fresh perspective when you’ve reached a dead end.” Generally, they valued its fast responses (ID 2_1, 24%): “It often saves me time that I would otherwise have to spend on lengthy Google searches.” In addition, ChatGPT was appreciated as a dialogue partner (ID 2_7, 12%), captured by responses such as “It can precisely explain the theoretical background and respond to follow-up questions” and “As a “contact person” for questions,” that offers a diversity of alternative explanations (ID 2_8, 10%), e.g., “providing explanations in other words.” Overall, participants saw an

TABLE IV. Categorization matrix for the fourth question (Q4: “Provide specific examples of how you use ChatGPT as a learning resource in physics.”), including anchor examples as well as the total number of respondents who were assigned the respective category. The number in the last column indicates the relative frequency with which the respective category has been assigned when excluding all participants who were assigned “not applicable.” It is important to note that some responses had multiple codes applied, leading to a percentage total greater than 100%.

ID	Code	tot #	rel # (%)
	<i>Example</i>		
1	Explanations on topics “I first have the content, formulas, and connections explained to me in detail before I can understand them at the level required at university.”	160	34
2	Starting point for problems “When I can’t find a starting point for certain problems, I ask ChatGPT and then see if the starting point makes sense and if I can work with it.”	97	21
3	Have problems solved “I let ChatGPT solve my exercises.”	72	15
4	Ask comprehension questions “If you have any questions about understanding.”	70	15
5	Create summaries “To summarize learning material and note the most important learning points.”	38	8
6	Definitions and technical terms “Querying definitions. Explaining terms. Finding formulas.”	26	6
7	Results check “To check my results and get a general overview of topics.”	38	8
8	Programming support “Programming in the master’s thesis.”	36	8
9	Research literature “I ask it for books that are suitable for my difficulties.”	26	6
10	Repetition and overview “ChatGPT provides good summaries of difficult material.”	30	6
11	Study plan and exam preparation “Create a study plan for exams; create flashcards.”	7	2
12	Discussion and sparring “I explain it to ChatGPT and ask it to point out any misunderstandings or things that sound wrong or inaccurate.”	10	2
13	Formulas and derivations “Formulas, derivations, integrals, practical examples, explanations.”	40	9
14	Examples and use cases “I also ask for examples of abstract theoretical concepts.”	6	1
15	Feedback and corrections “I let ChatGPT quiz me, and it tells me whether my answer was right or wrong and what I could improve.”	18	2
16	Creation of exercises “Furthermore, I also use AI as a learning partner to create practice exercises for me.”	6	1
17	Text editing and wording “I’m writing my BA thesis and I often ask whether certain phrases are correct or ask for brief explanations of things I don’t understand.”	11	2
18	Graphs and visualizations “For plotting graphs. Not much can go wrong and it’s easy to check.”	6	1

(Table continued)

TABLE IV. (Continued)

ID	Code	tot #	rel # (%)
	Example		
19	Comparison and alternatives “For example, if I need to compare two types of detectors in terms of different characteristics, I can ask ChatGPT for such a comparison.”	4	1
20	Quick information retrieval “Above all, it saves time. Instead of spending half an hour wading through various Wikipedia pages written in technical jargon, I have a simple summary that gives me an overview of the topic.”	36	8

advantage in the tool’s capacity for an individualized adaptation of the learning process (ID 2_2, 10%): “Can explain connections at an individually tailored level.” Moreover, few categories are concerned with affective learning variables, including motivation and frustration relief (ID 2_15, 2%), e.g., “Helps with approaches you wouldn’t otherwise have thought of and with tasks that leave you feeling desperate,” and “... helps you get out of a dead end, especially when you’re thinking in the wrong direction or don’t have a starting point at all,” as well as ChatGPT’s low threshold: “You can also ask questions without embarrassment that you would not otherwise dare to ask” (ID 2_3, 4%).

3. Results regarding the perceived risk of using ChatGPT for learning physics

Question 3 inverted question 2 by asking, “To what extent does the use of ChatGPT hinder the learning process in physics?” Here, only ten categories were extracted from the data, indicating greater convergence among participants’ perceptions (cf. Table III). By far, the most frequently emerging category is concerned with declining independence, as participants state, “ChatGPT tempts us to shorten the process of independent thinking,” undermining “... independent acquisition of knowledge,” thus leading to greater reliance on ChatGPT (ID 3_1, 39%). Another prominent category relates to incorrect or unreliable information (ID 3_3, 28%), where participants mention that “the results usually do not match” and “it can’t calculate and sometimes presents false claims as facts,” bridging the gap to question 1 (cf. IDs 1_2, 1_3, 1_6, and 1_10). Another hindrance was mentioned in the form of a lack of understanding when using ChatGPT for problem-solving (ID 3_2, 12%), as students noted that “you don’t have to actually engage with and understand the problem, which is why you barely learn anything new.” Several participants reported experiencing an illusion of understanding (ID 3_8, 5%), e.g., “It can give a false sense of security and distract from independent thinking” and “you become over-reliant on ChatGPT’s answers and only feel like you’re learning something,” perhaps stemming either from a superficial or abridged presentation

by ChatGPT (ID 3_5, 7%) or from uncritical acceptance of ChatGPT’s output (ID 3_4, 14%): “No independent thinking, just copying. Not much better than ‘copying from fellow students’.”

4. Results regarding specific examples of how ChatGPT is used as a learning resource in physics

Finally, question 4 asked participants to provide specific examples of how they use ChatGPT as a learning resource in physics. All 20 extracted categories are listed in Table IV. One third of all mentioned categories is concerned with using ChatGPT for explanations on topics (ID 4_1, 34%): “I first have the content, formulas, and connections explained to me in detail before I can understand them at the level required at university.” Similarly, ChatGPT is frequently used to provide an initial approach when solving complex problems (ID 4_2, 21%) to “help with exercises when I can’t figure them out on my own, thinking through the solution provided” or to “to answer questions or provide solutions to tasks,” reflecting the perception that ChatGPT is particularly useful for precisely this task (cf. ID 2_5). While it is also used as a tool for a simple results check (ID 4_7, 8%; “[Checking] the answers to exercises so that I know whether I have understood everything correctly”) or to quickly provide definitions of technical terms (ID 4_6, 6%; “When I don’t have specific definitions, I ask ChatGPT”), in more extreme cases, ChatGPT is used to solve homework or entire problem sheets (ID 4_3, 15%) as one participant stated, “I let ChatGPT solve my practice problems.” Furthermore, though more sparsely, the use cases comprise creative tasks, e.g., when using ChatGPT as a sparring partner (ID 4_12, 2%), e.g., “If I don’t understand a formula or would like to understand the derivation better, it’s quite nice to be able to ‘discuss’ it with ChatGPT,” make a study plan for exam preparation (ID 4_11, 2%; “Summary, explain topics, create learning plan”), ask comprehension questions (ID 4_4, 15%), e.g., “When I can’t get a question out of my head, I can’t search for it on the internet or wait until the next day to ask the professor, so I just ask GPT.” or even create new exercises (ID 4_16,

TABLE V. Overview of all model fit statistics for varying number of classes, including AIC, BIC, G^2 , sample-size adjusted BIC (ABIC), consistent AIC (CAIC), normalized entropy E, and class sizes. Lowest values are highlighted in bold font, with the exception of E, where the highest value is marked.

No. of classes	AIC	BIC	G^2	ABIC	CAIC	E	Class sizes (%)
1	17904	18096	5766	17973	18136	...	100
2	17219	17608	5000	17357	17687	0.804	30.2, 69.8
3	17262	17848	4963	17470	17967	0.872	0.6, 69.2, 30.2
4	17168	17951	4789	17446	18110	0.632	31.5, 15.2, 17.5, 35.8
5	17192	18172	4733	17541	18372	0.734	15.6, 27.0, 2.7, 31.7, 23.0

1%; “Creating sample exam questions”) and visualizations of physics problems (ID 4_18, 1%; “Envisioning through abstract images generated by ChatGPT.”) to “get an overview of topics.”

B. Results of the latent class analysis

1. Model selection

Model fit statistics for all models under investigation are provided in Table V and visualized in Fig. 2, ranging from 1 to 5 in terms of number of classes. As expected, the likelihood-ratio statistic G^2 decreases steadily with additional classes, reflecting improved fit at the cost of additional model complexity. Among the information criteria, the two-class solution provided the overall best balance between model fit and parsimony, as it exhibited the lowest values for BIC, ABIC, and CAIC, alongside a normalized entropy of 0.804, indicating high class discrimination. Although the three-class solution has a higher entropy, its fit indices increased slightly and the added class is of negligible size (0.6%), undermining practical usefulness. While the four- and five-class solutions produced lower

AIC and G^2 values, they are inferior in every other statistic—especially in the BIC, which is the most reliable fit statistic in LCA [99]. Taken together, the information criteria and entropy considerations point to the two-class model as providing the most adequate and interpretable representation of the data.

2. Class description

Estimated class population shares for the two-class solution are 69.8% for class 1 and 30.2% for class 2. Predicted class memberships based on the modal posterior probabilities were 65.0% and 35.0%, respectively, indicating a close correspondence between model estimates and individual classifications. The conditional item-response probabilities (i.e., the probability that a member of the respective class mentioned the respective category) for all selected variables are listed in Table VI. The two resulting class profiles are visualized in Fig. 3.

Overall, the class profiles are somewhat congruent with regard to question 1 but differ highly on all other questions. Members of class 1 most frequently mentioned categories 1_1 (Incorrect Calculations) and 1_7 (Errors in Specific or Complex Problems) in their responses to question 1; nonetheless, they proceeded to indicate endorsement of ChatGPT as a learning resource in physics, naming multiple instances of use cases. They generally view ChatGPT as helpful for learning physics, valuing it particularly for its quick responses (2_1), as well as its capability to promote understanding (2_4) and provide a starting point during problem-solving activities (2_5). More precisely, they use ChatGPT to provide initial explanations of various aspects when getting into a fresh topic (4_1), and they use it to solve homework or other exercises (4_3). On the other hand, class 2 exhibits slightly higher criticism toward ChatGPT, as indicated by higher response probabilities in question 1. The largest discrepancy is observed with regard to category 1_7 (Errors in Specific or Complex Problems), perhaps suggesting limited capabilities of ChatGPT to support advanced learners. Consequently, members of class 2 do not view ChatGPT as a helpful tool for physics learning, evidenced by item-response probabilities of 0 throughout all categories of

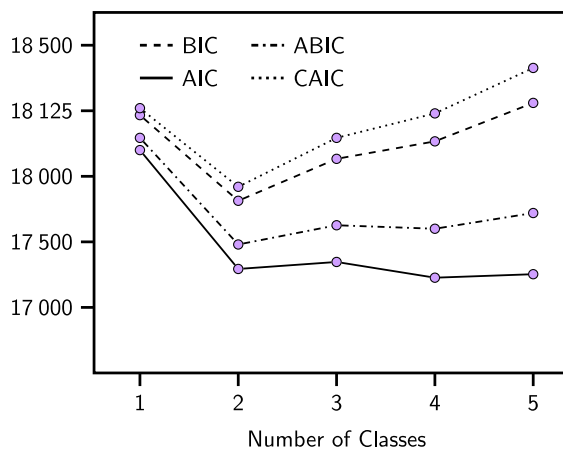


FIG. 2. All employed information-based criteria plotted against the number of classes. The visual “elbows” indicate that global minima are reached for the two-class solution in the case of BIC, ABIC, and CAIC. The AIC takes its lowest value for the four-class solution.

TABLE VI. Class-specific item-response probabilities for all selected variables.

Variable	Class 1	Class 2	Variable	Class 1	Class 2
1_1	27.9	24.2	3_1	0.0	23.7
1_2	11.3	12.3	3_2	0.0	7.2
1_3	8.5	10.3	3_3	0.0	16.9
1_4	5.3	9.2	3_4	0.0	8.5
1_5	3.8	5.8	3_5	0.0	4.2
1_7	16.1	24.5	3_6	0.0	5.2
1_9	4.1	2.5	3_7	0.0	4.6
1_12	0.6	14.2	3_8	0.0	2.9
2_1	20.1	0.6	3_10	0.0	3.9
2_2	8.7	0.0	4_1	22.1	1.0
2_4	29.7	0.0	4_2	12.8	1.9
2_5	21.8	0.0	4_3	10.1	0.0
2_6	10.0	0.0	4_4	9.5	0.8
2_7	10.7	0.0	4_5	5.3	0.0
2_8	8.2	0.0	4_6	3.2	1.1
2_9	5.1	0.0	4_7	5.2	0.4
2_11	5.3	0.0	4_8	4.6	1.0
2_12	6.9	0.0	4_9	3.5	0.4
...	...	...	4_10	3.9	0.8
...	...	...	4_13	5.6	0.0
...	...	...	4_20	4.8	0.6

question 2. Likewise, in their answers to question 4, they mentioned use cases only sparsely. In contrast to class 1, members of class 2 named several examples as to how ChatGPT can hinder learning physics, with 3_1 (Less Independence) and 3_3 (Incorrect or Unreliable Information) being the most frequently mentioned reasons.

Taken together, a coherent picture emerges from the LCA: Class 1 can be viewed as the user group that views ChatGPT as a beneficial tool for learning physics. Conversely, class 2 can be viewed as the nonuser group, expressed by a conservative and skeptical stance toward this technology. The sample of the study roughly splits into two thirds of class 1 (69.8%) and one third of class 2 (30.2%).

VI. DISCUSSION

This study’s central finding offers a unified answer to our research questions: at the current stage of adoption, the perception of ChatGPT among physics students (RQ1) is dichotomous, manifesting as two distinct user profiles (RQ2). The majority profile is the “Pragmatic User” (class 1; 69.8%), who utilizes the tool for learning support despite its known operational flaws. In contrast, the minority profile consists of “Skeptical Nonusers” (class 2; 30.2%). Their avoidance is driven by both the tool’s inaccuracies and a concern that it undermines the learning process by eroding independence and critical thinking. This user adoption rate is consistent with findings from higher education, where rates typically range from 61.4% to 84% [5–7,14]. More specifically, it aligns with a large-scale study in which 75% of physics students reported using the tool at least sometimes [32].

This dichotomy can be understood through the lens of technology acceptance, where students conduct a utility-risk analysis that currently seems to result in two divergent behaviors. Our findings add domain-specific detail to the “risk” component of this analysis: While current literature notes general student concerns about inaccuracy and error [12,31,73,76], our data reveal that for physics students, these concerns center on the tool’s inability to perform correct calculations and its difficulty in solving complex problems.

For the Pragmatic User, this utility-risk analysis resolves in favor of usefulness. This behavior is consistent with research identifying a positive relationship between students’ perceived risks and their intention to use ChatGPT [51,55]. The value students derive from the tool’s scaffolding—particularly for getting “unstuck”—is significant enough to outweigh these perceived risks, a calculation that may be supported by their self-efficacy in managing the technology’s limitations [64]. Their primary applications align with established uses, including seeking “further explanations” [12,69,71], clarifying

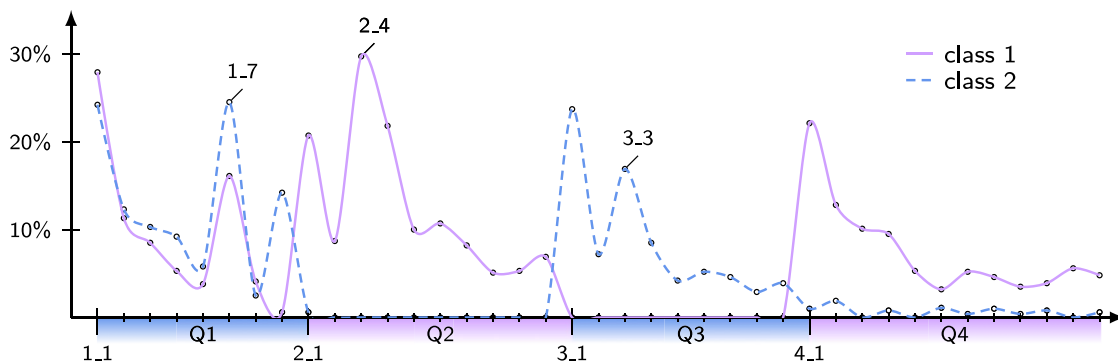


FIG. 3. Class-specific item-response probabilities for all selected variables. Respective sample shares are 69.8% for class 1 and 30.2% for class 2. Labels on the horizontal axis indicate the variables obtained from the categorization matrices, i.e., 1_1 is the category ID1 “Incorrect Calculations” from Q1. The horizontal axis is color coded to indicate which class predominates in each category.

“comprehension questions” [5,6,12,69,71], “having problems solved” [12], and performing “quick information retrieval” [5–7], e.g., for “formulas and derivations.” This emphasis on learning support diverges from other research where a majority of STEM students use AI for direct problem solving [12]. In our study, only a minority (15%) reported such use; their behaviors aligned more with obtaining procedural assistance, such as “promoting understanding” and finding “problem-solving starting points.”

In contrast, the utility-risk analysis for the Skeptical Nonuser leads to the opposite conclusion, a finding that adds nuance to recent research [51,55]. This prior research has operationalized risk primarily in terms of faulty output, academic dishonesty, and general “negative consequences” [51,55]. However, our data suggest that for nonusers, the perceived risks are fundamentally pedagogical, extending beyond operational concerns, mirroring the Pragmatic User’s awareness of technical flaws. Their dominant concerns center on the erosion of intellectual independence and the fostering of uncritical acceptance, aligning with previously documented concerns about overreliance and the potential for ChatGPT to undermine critical thinking skills [73,76].

The dichotomy between “Pragmatic user” and “skeptical nonuser” reflects differing strategies regarding the learning process. The pragmatic user’s reported approach aligns with instrumental help-seeking [100], using ChatGPT for support in the structured and metacognitive processes that constitute effective physics problem solving [101]. The actual effectiveness of these user-led strategies, however, has not yet been established. This risk lends credence to the Skeptical Nonuser’s position, which instead prioritizes the pedagogical necessity of productive failure. Research on productive failure consistently demonstrates that initial, unassisted effort on complex problems fosters more robust and transferable knowledge [102–104].

Our Pragmatic User and Skeptical Nonuser personas align with student archetypes identified in recent literature on AI adoption. For instance, the Pragmatic User persona encompasses the AI Advocates and Pragmatic Observers identified by Bewersdorff *et al.* [67], while the Skeptical Nonuser corresponds to their Cautious Critics. The Pragmatic User represents a variety of behaviors with varying levels of academic effectiveness. This group includes academically risky Assignment Delegators (23.1%); Knowledge Seekers, who use AI for information retrieval and conceptual understanding (16.4%); Proactive Learners, who engage in metacognitive strategies such as generating summaries and quizzes (11.8%); and All-rounders, who employ AI for all of these purposes [14]. In the context of problem solving, Russell *et al.* [72] similarly identified highly effective Selective Chatters, who achieve top exam scores with limited AI use, alongside the less successful heavier “chatters.” The Skeptic Nonuser comprises two distinct subgroups. The first subgroup of strategic nonusers is supported by evidence that top-performing students are overrepresented in low-AI-use

categories, such as versatile low reliers [14] and “high-performing nonchatters” [72]. This correlation suggests a rationale for this group’s approach. The second subgroup, identified as “low performers,” comprises struggling students who do not use it, even when it could be beneficial [72].

VII. LIMITATIONS

Beyond limitations inherent to a cross-sectional online questionnaire study (such as self-report bias, sampling bias, and temporal limitations), there are two important limitations that need to be considered to better contextualize the findings. First, the qualitative content analysis was largely conducted using a generative language model, with only partial human verification (10%) of both the codebook development and the annotation process. While this approach enabled the efficient handling of a large dataset, it may have introduced inaccuracies or overlooked nuances in participants’ responses. Therefore, the results should be interpreted with caution regarding potential biases in data coding and interpretation—the iterative refinement of the prompt addresses this limitation, but it should be considered regardless.

The other major limitation is related to the transformation of categories into categorical variables for the LCA: For example, a participant providing an answer to Q4 indicative of category 8 (“Programming Support”) and not mentioning anything else is assumed to not have used ChatGPT in any other scenario when learning physics. While it is not clear whether other use cases were forgotten by the participant or simply left out, it is coded precisely as a 4_8 in the LCA. In other words, the dataframe subjected to the LCA is likely to be heavily overpopulated with zeros, distorting class estimation. More simply, a zero in the obtained dataframe is to be understood as “not mentioned,” but is inherently coded as “no endorsement” in the LCA. Filtering out sparse variables and excluding participants who only named those variables mitigates this limitation but does not eliminate it.

VIII. RELEVANCE FOR EDUCATIONAL PRACTICE

Our identification of distinct user profiles provides a framework for a differentiated pedagogy that moves beyond one-size-fits-all approaches, fostering reflective AI use for pragmatic users while simultaneously validating the pedagogical approaches of skeptical nonusers.

Since 70% of students already use ChatGPT as a scaffold, instructors should proactively integrate the tool into lectures and seminars to guide, rather than ignore, its adoption, preventing AI from becoming a “crutch” that bypasses the learning process. Practical application should focus on fostering AI literacy through guided, reflective use. Assignments should require students to critically evaluate and justify the reasoning behind AI output.

Furthermore, instructors can co-create prompting templates together with their students, encouraging students to reflect on their own learning needs. For the minority group of Skeptical Nonusers, targeted exposure to the tool's potential in low-risk, bounded contexts (e.g., exploratory or feedback-oriented tasks) may help build trust and familiarity without compromising pedagogical integrity. It is important to note, however, that choosing not to use generative AI remains a valid stance, particularly for students who deliberately prioritize independent problem solving or productive struggle.

Overall, the goal is neither to enforce nor to discourage AI use but to cultivate critical, intentional, and discipline-specific engagement that leverages ChatGPT's benefits while preserving the intellectual rigor characteristic of physics learning.

IX. CONCLUSION AND OUTLOOK

This study provides one of the first large-scale qualitative analyses of how physics students engage with ChatGPT as a learning tool. Drawing on more than 1000 students' open-format responses and subsequent latent class analysis, two distinct user groups emerged. The majority group, labeled Pragmatic Users, employs ChatGPT strategically to enhance understanding, find starting points in problem solving, and access quick explanations, despite being aware of the model's inaccuracies. The minority group, Skeptical Nonusers, refrains from using ChatGPT, mostly due to concerns about the erosion of independent thinking and overreliance on this technology in general.

Our findings reveal that learners' adoption of generative AI is not uniform but shaped by their epistemic beliefs and perceived trade-offs between usefulness and risk. For many

students, ChatGPT functions as a scaffold that accelerates access to information and supports self-directed learning. For others, it threatens core pedagogical principles such as conceptual understanding and productive struggle.

Future research should explore how these user profiles evolve as generative AI becomes more deeply embedded in university teaching and as models gain improved reasoning capabilities. Longitudinal and mixed-methods studies could help determine whether pragmatic users develop sustained metacognitive strategies or drift into dependency. Conversely, future work should investigate whether Skeptical Nonusers adapt once institutional frameworks and AI-literacy programs increase trust and transparency. Additionally, it would be interesting to investigate whether students from each group show distinct technology acceptance or attitude profiles.

ACKNOWLEDGMENTS

This work was supported by the Open Access Publishing Fund of Leipzig University. This work is partially supported by the Ministry of Science and Culture of Lower Saxony through funds from the program *zukunft.niedersachsen* of the Volkswagen Foundation for the 'CAIMed – Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine' project (Grant No. ZN4257).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

-
- [1] F. M. Simon, R. K. Nielsen, and R. Fletcher, Generative AI and news report 2025: How people think about AI's role in journalism and society, Reuters Institute for the Study of Journalism, Oxford, Technical Report, 2025, [10.60625/risj-5bjv-yt69](https://www.risj.ox.ac.uk/publications/10.60625/risj-5bjv-yt69).
 - [2] Ofcom, Online nation 2024 report, Technical Report, Ofcom, 2024, <https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2024/online-nation-2024-report.pdf?v=386238>.
 - [3] University of South Carolina, College of Information and Communications, AI global public opinion tracker at USC: Wave 3 findings, august 2025, University of South Carolina, Columbia, SC, Technical Report, 2025, https://cms.sc.edu/study/colleges_schools/cic/docs/pdfs/ai_wave3.pdf.
 - [4] Initiative D21 e.V., D21-digital-index 2024/25: Jährliches Lagebild zur Digitalen Gesellschaft [D21 digital index 2024/25: annual report on the state of the digital society], Technical Report, Initiative D21 e.V., Berlin, 2025, https://initiatived21.de/uploads/03_Studien-Publikationen/D21-Digital-Index/2024-25/D21DigitalIndex_2024-2025_final.pdf.
 - [5] V. Morell-Mengual, O. Fernández-García, C. Berenguer, J. Ortega-Barón, M. D.;a D. Gil-Llario, and Vónica Estruch-García, Characteristics, motivations and attitudes of students using ChatGPT and other language model-based chatbots in higher education, *Educ. Inf. Technol.* **30**, 22257 (2025).
 - [6] Jörg von Garrel and J. Mayer, Artificial intelligence in studies—use of ChatGPT and AI-based tools among students in Germany, *Humanit. Soc. Sci. Commun.* **10**, 799 (2023).
 - [7] A. Schlude, U. Mendel, R. A. Stürz, and M. Fischer, Verbreitung und Akzeptanz generativer KI an Schulen und Hochschulen [Dissemination and acceptance of

- generative AI in schools and higher education institutions], report, Bayerisches Forschungsinstitut für digitale transformation (bidt) (2024).
- [8] J. Freeman, Student generative AI survey 2025, Higher Education Policy Institute (HEPI), Technical Report Policy Note 61, 2025, <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/>.
 - [9] S. Küchemann, A. Bauer, E. Becker, J. Veith, and P. Bitzenbauer, Zwischen aufbruch und skepsis—chancen und mögliche probleme bei der Nutzung von generativer KI im physikunterricht [Between optimism and skepticism—opportunities and potential challenges in the use of generative AI in physics education], *Naturwiss. Unterr. Phys.* **209**, 2 (2025).
 - [10] M. Smith, How are UK students really using AI? (2025), <https://yougov.co.uk/society/articles/52855-how-are-uk-students-really-using-ai>.
 - [11] Stack Overflow, 2025 developer survey: AI tools in the development process, Technical Report, Stack Overflow, 2025, https://survey.stackoverflow.co/2025/ai?utm_source=chatgpt.com.
 - [12] K. D. Wang, Z. Wu, L. I. Tufts, C. Wieman, S. Salehi, and N. Haber, Scaffold or crutch? Examining college students' use and views of generative AI tools for STEM education, *arXiv:2412.02653*.
 - [13] C. Köhler and J. Hartig, ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude, *Contemp. Educ. Technol.* **16**, ep528 (2024).
 - [14] A. Stojanov, Q. Liu, and J. H. L. Koh, University students' self-reported reliance on ChatGPT for learning: A latent profile analysis, *Comput. Educ.: Artif. Intell.* **6**, 100243 (2024).
 - [15] L. K. Dietrich and S. Grassini, Assessing ChatGPT acceptance and use in education: A comparative study among German-speaking students and teachers, *Educ. Inf. Technol.* **30**, 22151 (2025).
 - [16] A. Jabbar, S. U. Islam, and J. Boudjadar, A comparative review of LLM-based conversational systems: Insights from DeepSeek, ChatGPT, Gemini, Claude, and Copilot, in *Proceedings of the International Conference on AI and the Digital Economy (CADE 2025)* (IET, 2025), Vol. 2025, pp. 167–173.
 - [17] D. Ravšelj, D. Keržič, N. Tomažević, L. Umek, N. Brezovar, N. A. Iahad, A. A. Abdulla, A. Akopyan, M. W. Aldana Segura, J. AlHumaid *et al.*, Higher education students' perceptions of ChatGPT: A global study of early reactions, *PLoS One* **20**, e0315011 (2025).
 - [18] R. J. Eddine, E. Gide, and A. Al-Sabbagh, Generative AI in higher education: A cross-sector analysis of ChatGPT's impact on STEM, social sciences, and health-care, *STEM Educ.* **5**, 757 (2025).
 - [19] J. Sublime and I. Renna, Is ChatGPT massively used by students nowadays? A survey on the use of large language models such as ChatGPT in educational settings, *arXiv:2412.17486*.
 - [20] Y. Zhang, X. Xie, and Q. Xu, ChatGPT in medical education: Bibliometric and visual analysis, *JMIR Med. Educ.* **11**, e72356 (2025).
 - [21] S. Z. S. Abdalla, Understanding ChatGPT adoption for data analytics learning: A UTAUT perspective among social science students in Oman, *Social Sci. Humanit. Open* **11**, 101310 (2025).
 - [22] M. Ma, Exploring the acceptance of generative artificial intelligence for language learning among EFL postgraduate students: An extended TAM approach, *Int. J. Appl. Linguist.* **35**, 91 (2025).
 - [23] Q.-u.-A. Mirza and N. S. Rebello, Help or hype? Students' engagement and perception of using AI to solve physics problems, *arXiv:2509.25642*.
 - [24] P. Wattanakasiwich, K. Kaewkhong, and D. Katwibun, Physics instructors' acceptance and implementation of generative AI, *Phys. Rev. Phys. Educ. Res.* **21**, 010155 (2025).
 - [25] L. Krupp, S. Steinert, M. Kiefer-Emmanouilidis, K. E. Avila, P. Lukowicz, J. Kuhn, S. Kuchemann, and J. Karolus, Unreflected acceptance: Investigating the negative consequences of ChatGPT-assisted problem solving in physics education, in *Proceedings of the HHAI 2024: Hybrid Human AI Systems for the Social Good* (2024), 10.3233/FAIA240195.
 - [26] D. Tong, Y. Tao, K. Zhang, X. Dong, Y. Hu, S. Pan, and Q. Liu, Investigating ChatGPT-4's performance in solving physics problems and its potential implications for education, *Asia Pacific Educ. Rev.* **25**, 1379 (2024).
 - [27] K. D. Wang, E. Burkholder, C. Wieman, S. Salehi, and N. Haber, Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving, *Front. Educ.* **8**, 1330486 (2024).
 - [28] R. Horchani, ChatGPT's problem-solving abilities in context-rich and traditional physics problems, *Phys. Educ.* **60**, 025019 (2025).
 - [29] V. D.:ctor López-Simó and M. F. Rezende Jr., Challenging ChatGPT with different types of physics education questions, *Phys. Teach.* **62**, 290 (2024).
 - [30] G. Polverini, J. Melin, E. Önerud, and B. Gregorcic, Performance of ChatGPT on tasks involving physics visual representations: The case of the brief electricity and magnetism assessment, *Phys. Rev. Phys. Educ. Res.* **21**, 010154 (2025).
 - [31] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT, *Phys. Rev. Phys. Educ. Res.* **19**, 010142 (2023).
 - [32] B. Agyare, J. Asare, A. Kraishan, I. Nkrumah, and D. K. Adjekum, A cross-national assessment of artificial intelligence (AI) chatbot user perceptions in collegiate physics education, *Comput. Educ.* **8**, 100365 (2025).
 - [33] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Q.* **13**, 319 (1989).
 - [34] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, User acceptance of information technology: Toward a unified view, *Manage. Inf. Syst. Q.* **27**, 425 (2003).
 - [35] H. M. Selim, An empirical investigation of student acceptance of course websites, *Comput. Educ.* **40**, 343 (2003).
 - [36] Bštjan Šumak, M. Heričko, and M. Pušnik, A meta-analysis of e-learning technology acceptance: The role of

- user types and e-learning technology types, *Comput. Hum. Behav.* **27**, 2067 (2011).
- [37] K. A. Pituch and Y.-K. Lee, The influence of system characteristics on e-learning use, *Comput. Educ.* **47**, 222 (2006).
- [38] C.-M. Chiu and E. T. G. Wang, Understanding web-based learning continuance intention: The role of subjective task value, *Inf. Manage.* **45**, 194 (2008).
- [39] L. Xue, A. M. Rashid, and S. Ouyang, The unified theory of acceptance and use of technology (UTAUT) in higher education: A systematic review, *Sage Open* **14**, 21582440241229570 (2024).
- [40] F. D. Davis. A technology acceptance model for empirically testing new end-user information systems: Theory and results, Ph.D. thesis, Massachusetts Institute of Technology, 1985.
- [41] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, User acceptance of computer technology: A comparison of two theoretical models, *Manage. Sci.* **35**, 982 (1989).
- [42] F. D. Davis and A. Granić, *The Technology Acceptance Model: 30 Years of TAM*, Springerbriefs in Human-Computer Interaction Series (Springer, Cham, 2024), 10.1007/978-3-030-45274-2.
- [43] W. R. King and J. He, A meta-analysis of the technology acceptance model, *Inf. Manage.* **43**, 740 (2006).
- [44] M. Blut, A. Y. L. Chong, Z. Tsiga, and V. Venkatesh, Meta-analysis of the unified theory of acceptance and use of technology (UTAUT): Challenging its validity and charting a research agenda in the red ocean, *J. Assoc. Inf. Syst.* **23**, 13 (2022).
- [45] S. Taylor and P. Todd, Assessing IT usage: The role of prior experience, *MIS Q.* **561** (1995).
- [46] V. Venkatesh and F. D. Davis, A theoretical extension of the technology acceptance model: Four longitudinal field studies, *Manage. Sci.* **46**, 186 (2000).
- [47] Y. Wei, C. Wang, S. Zhu, H. Xue, and F. Chen, Online purchase intention of fruits: Antecedents in an integrated model based on technology acceptance model and perceived risk theory, *Front. Psychol.* **9**, 1521 (2018).
- [48] P. A. Pavlou, Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model, *Int. J. Electron. Commerce* **7**, 101 (2003).
- [49] V. Venkatesh, J. Y. L. Thong, and X. Xu, Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology, *Manage. Inf. Syst. Q.* **36**, 157 (2012).
- [50] M. Limayem, S. G. Hirt, and C. M. K. Cheung, How habit limits the predictive power of intention: The case of information systems continuance, *Manage. Inf. Syst. Q.* **31**, 705 (2007).
- [51] A. Shuhaiber, M. A. Kuhail, and S. Salman, ChatGPT in higher education: A student's perspective, *Comput. Human Behav. Rep.* **17**, 100565 (2025).
- [52] M. Fetaji, Devising a model AI-UTAUT by combining artificial intelligence AI with unified theory of acceptance and use of technology (UTAUT), *SAR J. Sci. Res.* **6**, 182 (2023).
- [53] I. Baroni, G. Re Calegari, D. Scandolari, and I. Celino, AI-TAM: A model to investigate user acceptance and collaborative intention in human-in-the-loop AI applications, *Hum. Comput.* **9**, 1 (2022).
- [54] S. Grassini, M. L. Aasen, and A. Møgelvang, Understanding university students' acceptance of ChatGPT: Insights from the UTAUT2 model, *Appl. Artif. Intell.* **38**, 2371168 (2024).
- [55] S. Balaskas, V. Tsiantos, S. Chatzifotiou, and M. Rigou, Determinants of ChatGPT adoption intention in higher education: Expanding on TAM with the mediating roles of trust and risk, *Information* **16**, 82 (2025).
- [56] M. Sallam, N. A. Salim, M. Barakat, K. Al-Mahzoum, A'a B. Al-Tammemi, D. Malaeb, R. Hallit, and S. Hallit, Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study, *JMIR Med. Educ.* **9**, e48254 (2023).
- [57] C. Y. Lai, K. Y. Cheung, C. S. Chan, and K. K. Law, Integrating the adapted UTAUT model with moral obligation, trust and perceived risk to predict ChatGPT adoption for assessment support: A survey with students, *Comput. Educ.* **6**, 100246 (2024).
- [58] P. Bhaskar, P. Misra, and G. Chopra, Shall i use ChatGPT? A study on perceived trust and perceived risk towards ChatGPT usage by teachers at higher education institutions, *Int. J. Inf. Learn. Technol.* **41**, 428 (2024).
- [59] A. Strzelecki, To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology, *Interact. Learn. Environ.* **32**, 5142 (2023).
- [60] M. Abdaljaleel, M. Barakat, M. Alsanafi *et al.*, A multinational study on the factors influencing university students' attitudes and usage of ChatGPT, *Sci. Rep.* **14**, 1983 (2024).
- [61] K. Parveen, T. Q. B. Phuc, A. A. Alghamdi *et al.*, Unraveling the dynamics of ChatGPT adoption and utilization through structural equation modeling, *Sci. Rep.* **14**, 23469 (2024).
- [62] A. E. E. Sobaih, I. A. Elshaer, and A. M. Hasanein, Examining students' acceptance and use of ChatGPT in Saudi Arabian higher education, *Eur. J. Investig. Health Psychol. Educ.* **14**, 709 (2024).
- [63] H. Albayati, Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study, *Comput. Educ.* **6**, 100203 (2024).
- [64] M. F. Shahzad, S. Xu, and M. Asif, Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model, *Br. Educ. Res. J.* **51**, 489 (2025).
- [65] A. J. S. Morin, A. Bujacz, and Mène Gagné, Person-centered methodologies in the organizational sciences: Introduction to the feature topic, *Organ. Res. Methods* **21**, 803 (2018).
- [66] H. Chang, B. Liu, Y. Zhao, Y. Li, and F. He, Research on the acceptance of ChatGPT among different college student groups based on latent class analysis, *Interact. Learn. Environ.* **33**, 22 (2024).
- [67] A. Bewersdorff, M. Hornberger, C. Nerdel, and D. S. Schiff, AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy, *Comput. Educ.* **8**, 100340 (2025).

- [68] G. Fenu, R. Galici, M. Marras, and D. Reforgiato, Exploring student interactions with AI in programming training, in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, UMAP Adjunct '24* (, Cagliari, Italy, 2024) (Association for Computing Machinery, Cagliari, Italy, 2024), pp. 555–560, [10.1145/3631700.3665227](https://doi.org/10.1145/3631700.3665227).
- [69] R. Arum, M. Calderon Leon, X.F. Li, and J. Lopes, ChatGPT early adoption in higher education: Variation in student usage, instructional support, and educational equity, *AERA Open* **11**, 23328584251331956 (2025).
- [70] H. Yilmaz, S. Maxutov, A. Baitekov, and N. Balta, Students' perception of ChatGPT: A technology acceptance model study, *Int. Educ. Rev.* **1**, 57 (2023).
- [71] F. Valeri, P. Nilsson, and A.-M. Cederqvist, Exploring students' experience of ChatGPT in STEM education, *Comput. Educ.* **8**, 100360 (2025).
- [72] J.-E. Russell, A. M. Smith, S. George, J. Pratt, B. Fodale, C. Monk, and A. Brummett, Unlocking insights: Investigating student AI tutor interactions in a large introductory STEM course, in *Proceedings of the 15th International Learning Analytics and Knowledge Conference, LAK '25* (Association for Computing Machinery, New York, NY, 2025), pp. 451–461, [10.1145/3706468.3706524](https://doi.org/10.1145/3706468.3706524).
- [73] T. Kregear, M. Babayeva, and R. Widenhorn, Analysis of student interactions with a large language model in an introductory physics lab setting, *Int. J. Artif. Intell. Educ.* **35**, 2993 (2025).
- [74] J. Wang and W. Fan, The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis, *Humanit. Soc. Sci. Commun.* **12**, 621 (2025).
- [75] T. El Fathi, A. Saad, H. Larhzil, D. Lamri, and E. M. Al Ibrahim, Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance, *Discip. Interdiscip. Sci. Educ. Res.* **7**, 6 (2025).
- [76] A. Lieb and T. Goel, Student interaction with newtbot: An LLM-as-tutor chatbot for secondary physics education, in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24* (Association for Computing Machinery, Honolulu, HI, 2024), p. 614, [10.1145/3613905.3647957](https://doi.org/10.1145/3613905.3647957).
- [77] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, GPT-4 Technical Report, [arXiv:2303.08774](https://arxiv.org/abs/2303.08774).
- [78] Z. O. Dunivin, Scalable qualitative coding with llms: Chain-of-thought reasoning matches human performance in some hermeneutic tasks, [arXiv:2401.15170](https://arxiv.org/abs/2401.15170).
- [79] R. Bijker, S. S. Merkouris, N. A. Dowling, and S. N. Rodda, ChatGPT for automated qualitative research: Content analysis, *J. Med. Internet Res.* **26**, e59050 (2024).
- [80] S. B. Merriam, Qualitative research: Designing, implementing, and publishing a study, in *Handbook of Research on Scholarly Publishing and Research Methods* (IGI Global Scientific Publishing, San Francisco, CA, 2015), pp. 125–140.
- [81] Y. Intrator, M. Halfon, R. Goldenberg, R. Tsarfaty, M. Eyal, E. Rivlin, Y. Matias, and N. Aizenberg, Breaking the language barrier: Can direct inference outperform pre-translation in multilingual LLM applications?, in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, edited by K. Duh, H. Gomez, and S. Bethard (Association for Computational Linguistics, Mexico City, Mexico, 2024), pp. 829–844, [10.18653/v1/2024.naacl-short.75](https://doi.org/10.18653/v1/2024.naacl-short.75).
- [82] <https://github.com/JohannaSchrader/ChatGPT-Learners-Qualitative>.
- [83] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, The curious case of neural text degeneration, *Proceedings of the International Conference on Learning Representations (ICLR)* (2020).
- [84] G. Hinton, O. Vinyals, and J. Dean, Distilling the knowledge in a neural network, in *NIPS Deep Learning and Representation Learning Workshop* (2015), [10.48550/arXiv.1503.02531](https://arxiv.org/abs/1503.02531).
- [85] L. Li, S. Geng, Z. Li, Y. He, H. Yu, Z. Hua, G. Ning, S. Wang, T. Xie, and H. Yang, Infibench: Evaluating the question-answering capabilities of code large language models, *Adv. Neural Inf. Process. Syst.* **37**, 128668 (2024).
- [86] OpenAI, API reference—OpenAI platform.
- [87] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, Lost in the middle: How language models use long contexts, *Trans Assoc. Comput. Linguist.* **12**, 157 (2024).
- [88] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, Language models are few-shot learners, *Adv. Neural Inf. Process. Syst.* **33**, 1877 (2020), [arXiv:2005.14165](https://arxiv.org/abs/2005.14165).
- [89] P. Törnberg, Best practices for text annotation with large language models, *Sociologica* **18**, 67 (2024).
- [90] E. Machisi, A. Kurudirek, and J. M. Veith, Examining the least preferred topics among advanced placement calculus student, *Eurasia J. Math. Sci. Technol. Educ.* **21**, em2712 (2025).
- [91] L. M. Collins and S. T. Lanza, *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences* (John Wiley & Sons, 2009), [10.1002/9780470567333](https://doi.org/10.1002/9780470567333).
- [92] M. Wang, M. Sundstrom, K. Nylund-Gibson, and M. Ing, Advancing clustering methods in physics education research: A case for mixture models, *Phys. Rev. Phys. Educ. Res.* **21**, 020126 (2025).
- [93] B. E. Weller, N. K. Bowen, and S. J. Faubert, Latent class analysis: A guide to best practice, *J. Black Psychol.* **46**, 287 (2020).
- [94] E. Palmgren, K. Tuominen, and I. Kontro, Self-efficacy and conceptual knowledge in quantum mechanics during teaching reforms and the COVID-19 pandemic, *Phys. Rev. Phys. Educ. Res.* **18**, 020122 (2022).
- [95] H. Bekaert, H. Van Winckel, W. Van Dooren, A. Steegen, and M. De Cock, Identifying students' mental models of the apparent motion of the sun and stars, *Phys. Rev. Phys. Educ. Res.* **18**, 010130 (2022).

- [96] C. J. Van Lissa, M. Garnier-Villarreal, and D. Anadria, Recommended practices in latent class analysis using the open-source R-package tidy, *SEM: Struct. Equation Model.* **31**, 526 (2024).
- [97] J.-Y. Tein, S. Coxe, and H. Cham, Statistical power to detect the correct number of classes in latent profile analysis, *Struct. Equation Model.* **20**, 640 (2013).
- [98] D. A. Linzer and J. B. Lewis, Polca: An R package for polytomous variable latent class analysis, *J. Stat. Softw.* **42**, 1 (2011).
- [99] O. Lezhnina and Gábor Kismihók, Latent class cluster analysis: Selecting the number of clusters, *MethodsX* **9**, 101747 (2022).
- [100] S. A. Karabenick and J. R. Knapp, Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students, *J. Educ. Psychol.* **83**, 221 (1991).
- [101] T. Musengimana, L. L. Yadav, J. Uwamahoro, and G. Nizeyimana, Instructional strategies for enhancing students' problem-solving skills in physics: A systematic review, *Discover Educ.* **4**, 380 (2025).
- [102] M. Kapur, Productive failure, *Cognit. Instr.* **26**, 379 (2008).
- [103] A. Jackson, A. Godwin, S. Bartholomew, and N. Mentzer, Learning from failure: A systematized review, *Int. J. Technol. Design Educ.* **32**, 1853 (2022).
- [104] S. A. Pathak, B. Kim, M. J. Jacobson, and B. Zhang, Learning the physics of electricity: A qualitative analysis of collaborative processes involved in productive failure, *Int. J. Comput. Supported Collab. Learn.* **6**, 57 (2011).