

ORIGINAL ARTICLE

Exploring the Application of Large Language Models in Instrumentation and Control Engineering Education: A Questionnaire Survey and Examination Performance Analysis

Wu Xu  | Zhang Wei | Peng Yan

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, People's Republic of China

Correspondence: Peng Yan (py@usst.edu.cn)

Received: 20 July 2024 | **Revised:** 1 January 2025 | **Accepted:** 2 January 2025

Funding: This work was supported by Shanghai Municipal Education Commission Key Course Construction Project.

Keywords: ChatGPT | engineering education | Instrumentation and Control Engineering | Large Language Models | questionnaire

ABSTRACT

This study investigates the use of Large Language Models (LLMs) by undergraduates majoring in Instrumentation and Control Engineering (ICE) at University of Shanghai for Science and Technology. We conducted a questionnaire survey to assess the awareness and usage habits of these LLMs among ICE undergraduates in ICE courses, focusing on the model types, frequency, self-study tasks and post-usage feedback. The survey results indicate that ChatGPT is the most frequently used model among ICE undergraduates. Furthermore, we evaluated the performance of GPT-3.5 and GPT-4 on a set of 342 ICE questions sourced from the China MOOC platform. The results reveal significant performance differences between GPT-3.5 and GPT-4 across various ICE sub-disciplines and question types. GPT-4 shows particular proficiency in understanding control technology and system, achieving an accuracy of 82.16%, compared to 68.42% for GPT-3.5. Finally, we explore the factors influencing students use LLMs for ICE self-study and propose strategies for teachers to incorporate LLMs into ICE courses. This study highlights the potential of LLMs to enhance ICE education and provides concrete examples of their application in engineering coursework.

1 | Introduction

With the rapid advancement of artificial intelligence, Large Language Models (LLMs) have emerged as a major innovation in Natural Language Processing (NLP) (Karanikolas et al. 2024; Chen et al. 2024). Since the introduction of the Transformer architecture in 2017 (Vaswani et al. 2017), several LLMs based on this architecture have been released, greatly enhancing the efficiency and capability of NLP tasks (Ansar, Goswami, and Chakrabarti 2024; Kheddar 2024). Between 2018 and 2019, Google and its collaborators launched BERT, XLNet and T5 (Text-To-Text Transfer Transformer), achieving breakthroughs in language understanding and task transferability (Zhilin Yang et al. 2019; Jacob Devlin, Lee, and Toutanova 2019; Kale and Rastogi 2020). In 2019, Baidu

developed the ERNIE model, which improved Chinese language comprehension through the integration of structured knowledge and multimodal learning (Sun et al. 2020, 2021). In 2022, OpenAI released ChatGPT, based on the GPT-3.5 and GPT-4 models, excelling in traditional NLP tasks such as machine translation and text summarization (Hendy et al. 2023; Goyal, Li, and Durrett 2022), as well as emerging applications like conversation systems and writing assistants (Imran and Almusharraf 2023). ChatGPT marked the beginning of large-scale public deployment of LLMs (Haleem, Javaid, and Singh 2022). More recently, Anthropic's Claude model and Meta AI's LLaMA model, both based on GPT architectures, have developed. The Claude model enhances conversational generation and task adaptability by incorporating human feedback during training (Adetayo, Aborisade, and Sanni 2024).

The growing use of LLMs in education is leading to a transition from traditional teaching styles to more customised, learner-focused approaches (Wang et al. 2024; Javaid et al. 2023; Jeon and Lee 2023). Firstly, LLMs are effective for explaining knowledge (Fitria 2023; Li, Carlson, and Potts 2022). For instance, XLNet and BERT are particularly good at text analysis and extracting information, which helps with classification and concept understanding (Somers, Cunningham-Nelson, and Boles 2021). ERNIE, aimed at Chinese content, boosts comprehension of complex contexts by using knowledge graphs (Sun et al. 2019). In addition, LLMs can enhance personalised education by establishing objectives, outlining routes and suggesting materials. For instance, ChatGPT tailors educational pathways according to student needs and performance (Abas et al. 2023; Limo et al. 2023), supplying coherent answers that facilitate problem-solving and assignment completion (Moore et al. 2022). This ensures timely and effective guidance. LLMs also help teachers with educational data analysis and decision-making (Ghavidel, Zouaq, and Desmarais 2020). For instance, BERT is able to process large amounts of student feedback and provide sentiment analysis (Zhang 2022; Alhazmi, Khan, and Syed 2023). This allows teachers to gain insights into student emotions and optimise their teaching strategies. The existing literature suggests that assessing the accuracy of LLMs through specialised simulated exams provides direct evidence of their applicability in specific educational fields (Xu and Ouyang 2022; Newton and Xiromeriti 2024). For instance, the examination performance of ChatGPT in fields such as English (de Winter 2023), computer science (Finnie-Ansley et al. 2022; Kosar et al. 2024), medicine (Ali et al. 2023; Angel et al. 2022; Sadeq et al. 2024), and law (Freitas and Gomes 2023; Chalkidis 2023; Katz et al. 2024), has been analysed.

Research on the application of LLMs in engineering education, especially in Instrumentation and Control Engineering (ICE), is still in its nascent stages. ICE, a foundational discipline in automation and intelligent manufacturing, necessitates that students master both complex engineering principles and practical competencies. The complexity, interdisciplinary scope and practical orientation of ICE courses, however, present significant challenges, particularly in areas such as mathematical modelling and control system design. Notably, some students have begun independently using LLMs to support self-study, even in the absence of teacher guidance. This suggests both the potential value of LLMs in education and an increasing ‘awareness’ among students. Despite these developments, there is currently no research about the applications of LLMs in ICE education or assessing their performance in ICE simulated exam.

This study investigates the application of LLMs in ICE self-study among undergraduates at the University of Shanghai for Science and Technology. Through a questionnaire survey, we assess their ‘awareness’ and ‘usage habits’ regarding LLMs as ICE self-study tools. ‘Awareness’ refers to their familiarity with the existence and functions of LLMs (Shahzad, Xu, and Javed 2024), while ‘usage habits’ refers to their behavioural patterns in using these tools for specific tasks. The survey focuses on the types of LLMs used (e.g. ChatGPT, Claude, LLaMA, Ernie Bot, Tongyi), their usage frequency, purposes in self-study and feedback on usage. Additionally, we evaluate the performance of the most popular models (based on survey responses) in a simulated ICE exam, to assess their accuracy and domain-specific expertise. Finally, we

identify factors influencing LLM use in ICE self-study and provide recommendations to enhance LLM integration in ICE courses.

2 | Methods

2.1 | Questionnaire Design

The questionnaire comprises both objective and subjective questions, totaling 17 questions (Q1–Q17, sequentially numbered). It covers four dimensions: basic information of respondents, awareness of LLMs in education, usage habits of LLMs in ICE self-study, and post-usage feedback about LLMs in ICE self-study (Table 1).

2.2 | Online Questionnaire Survey

The survey participants are undergraduate students majoring in ICE at the University of Shanghai for Science and Technology (USST). All participants have completed courses in the Principles of Automatic Control and possess foundational knowledge of control theory.

This study utilised a real-name questionnaire survey, preceded by pilot tests with 30 participants to address any ‘expressed deviation’ (May–June, 2024). Adjustments were made to the number and difficulty of questions, with an excepted completion time of 10–15 min. The questionnaire was designed as a single scrolling page, including an introduction that explain the purpose and significance of the survey.

The questionnaire was distributed via the USST self-compiled Education Online platform (<https://1906home.usst.edu.cn/>), and shared with students through mass email. The survey was conducted from October 8 to October 19, 2024, yielding 98 valid responses. Data collection and analysis were performed using the Education Online platform, which generated statistical

TABLE 1 | Four comprises of questionnaire.

Dimension 1	Basic information of respondents, including name, gender, age, education and academic background, E-mail, a total of seven problems
Dimension 2	Awareness of LLMs in education, encompassing their understanding of LLMs, familiarity with specific models, and awareness of real-world applications of LLMs in education (Table 2)
Dimension 3	Usage habits of LLMs in ICE self-study, covering independent use in ICE coursework, types of LLMs applied, tasks completed, and frequency of usage (Table 3)
Dimension 4	Post-usage feedback about LLMs in ICE self-study, including perceived benefits, encountered challenges, and evaluations of academic suitability (Table 4)

TABLE 2 | Awareness analysis of LLMs in education among respondents.

Question	Options	Number of Votes	Percentage
Do you understand the concept of large language models (LLMs)? (Q8: Single-choice question)	Yes, I understand	97	99.0%
	No, I do not understand	0	0%
	Not sure	1	1.0%
Have you seen or read about real-world applications of LLMs in education? (Q9: Single-choice question)	Yes, I have	97	99.0%
	No, I haven't	0	0%
	Not sure	1	1.0%
Which specific LLM tools are you familiar with? (Q10: Multiple-choice question)	ChatGPT (by OpenAI)	98	100.0%
	Claude (by Anthropic)	2	2.0%
	LLaMA (by Meta)	1	1.0%
	Ernie Bot (by Baidu)	56	57.1%
	iFLYTEK Spark (iFLYTEK)	19	19.4%
	Tongyi (by Alibaba)	33	33.7%
	None of these	0	0
	Others (please specify)	9	9.2%

results while confidentiality of respondents by excluding names and email addresses.

2.3 | Question Bank for ICE Simulated Exam

Studies show that specialised exam on LLMs demonstrate their educational value in specific fields. This study focuses on the most frequently used LLMs identified in the student questionnaire, specifically assessing their performance in ICE exams. The questions for the simulated ICE exam were sourced from various ICE courses available on the China MOOC platform, taught by faculty teams from eight Chinese universities recognised for their high standing in the ICE discipline. These courses provide standardised exam questions and reference answers, which constitute the question bank for the LLM assessments. The question bank encompasses eight sub-specialties within ICE. Visual content, such as industrial control system diagrams and control scheme block diagrams, was excluded due to the inability of current LLMs to process such information. Ultimately, the final question bank comprises 342 questions presented in their original Chinese form.

2.4 | Performance Assessment for LLMs on ICE Simulated Exam

In our student survey, ChatGPT was the most frequently used LLM among students. Here we assess the performance of GPT-3.5 and GPT-4 models on an ICE simulated exam. Assessment for GPT-3.5 and GPT-4 models were conducted on June 16–18 and June 15–17, 2024, respectively. Questions were manually input into the ChatGPT interface, and model responses were compared with reference answers from MOOC courses. The assessment focused on model accuracy across various sub-specialties and question types within ICE.

3 | Results

3.1 | Questionnaire Analysis of LLMs in ICE Self-Study

3.1.1 | Basic Information of the Respondents

We collected 98 valid questionnaires from undergraduates majoring in ICE. The group was 84% male and 16% female. The age distribution was as follows: 35.7% were 21 years old (mostly third-year students), 33.6% were 22 years old (mostly fourth-year students) and 30.5% were 20 years old (mostly second-year students). All respondents had completed key courses of the ICE curriculum like automatic control principles, engineering measurement techniques, and relevant laboratory courses.

3.1.2 | Awareness of LLMs in Education Among Respondents

Dimension 2 (Q8–Q10) assessed respondents' awareness of LLMs, including their understanding of basic concepts, familiarity with specific models, and knowledge of educational applications. To ensure data accuracy, the questionnaire included a 'Not sure' option for respondents with uncertain views.

Responses to Q8 show that 99.0% of respondents understood the basic concept of LLMs, indicating widespread recognition and growing interest in artificial intelligence among students. These may pave the way for future LLM applications in ICE self-study. Responses to Q9 show indicate that 99.0% of respondents had encountered LLM applications in education. Interviews reveal that most students learned about

these applications through online forums and social media. Responses to Q10 reveal how familiarity respondents with specific models. ChatGPT (OpenAI's GPT models) was known to 100.0% of respondents. This was followed by Ernie Bot (Baidu's ERNIE model), known to 57.1% of respondents, and Tongyi (Alibaba's Qwen-7B model), known to 33.7%. Other LLMs, such as Claude, LLaMA, and iFLYTEK Spark, also had some familiarity among respondents.

Overall, these results indicate high recognition of LLMs among respondents. However, there were notable differences in familiarity across specific models. ChatGPT held the leading popularity, likely due to its extensive global usage. Respondents also showed greater familiarity with Chinese-developed LLMs, such as Ernie Bot and Tongyi, which receive more promotion on Chinese social media. In contrast, limited promotion of foreign models like Claude and LLaMA on Chinese social media appears to reduce their visibility, thereby limiting their popularity in China.

3.1.3 | Usage Habits of LLMs in ICE Self-Study Among Respondents

Dimension 3 (Q11–Q14) investigated how respondents use LLMs in ICE courses, focusing on focusing on the types of models employed, tasks undertaken and frequency of use. The results are summarised in Table 3.

Response to Q11 show that 83.7% of respondents have used LLMs for ICE self-study, compared to 16.3% for other courses. This suggests a widespread acceptance of LLMs in ICE. Regarding model types, Q12 responses reveal a clear preference for specific LLMs among respondents. ChatGPT was the most frequently used model (72.4%), followed by Baidu's Ernie Bot (50.0%), Alibaba's Tongyi (28.6%), and iFLYTEK Spark (17.3%). Foreign models, like Anthropic's Claude and Meta's LLaMA, were less frequently chosen. The preference for specific models is likely attributable to their accessibility, affordability, localised support and marketing efforts. In contrast, foreign models face

TABLE 3 | Survey data on LLMs usage for ICE self-study.

Question	Options	Number of votes	Percentage
Have you independently used LLMs in ICE-related coursework without teacher guidance? (Q11: Single-choice question)	Yes, I have	82	83.7%
	No, I have only used LLMs in other coursework	16	16.3%
	I have not used LLMs in my coursework	0	0.0%
What types of LLMs have you applied in your ICE-related coursework? (Q12: Multiple-choice question)	ChatGPT (by OpenAI)	71	72.4%
	Claude (by Anthropic)	0	0.0%
	LLaMA (by Meta)	0	0.0%
	Ernie Bot (by Baidu)	49	50.0%
	iFLYTEK Spark (iFLYTEK)	17	17.3%
	Tongyi (by Alibaba)	28	28.6%
	None of these	0	0.0%
	Others (please specify)	9	9.2%
What specific tasks have you completed using LLMs in ICE-related coursework? (Q13: Multiple-choice question)	Problem-solving	57	58.2%
	Homework assistance	68	69.4%
	Project and report generation	38	38.8%
	Simulation Coding	21	21.4%
	Data Analysis and Simulation	19	19.4%
	Exam Preparation	4	4.1%
	None of these	0	0.0%
	Others (please specify)	0	0.0%
How frequently do you utilise LLMs in ICE-related coursework? (Q14: Single-choice question)	Daily (at least 4–5 times per week)	5	5.1%
	Weekly (1–3 times per week)	31	31.6%
	Monthly (about 1–3 times per month)	39	39.8%
	Rarely (less than once per month)	7	7.1%
	Never	20	20.4%

challenges due to restricted access, high costs and less promotional activities in China.

Responses to Q13 indicate that LLMs are primarily used for homework assistance (69.4%), problem-solving (58.2%), and project/report generation (38.8%). Fewer respondents use LLMs for simulation coding (21.4%), data analysis (19.4%), and exam preparation (4.1%). These results suggest that LLMs are more commonly applied to general tasks, with less frequent use in specialised tasks.

Q14 investigates the frequency of LLM use among ICE students. Among the 82 respondents who used LLMs for self-study, the frequency ranged from daily to rarely. Specifically, 5 respondents used LLMs 4–5 times per week (daily), 31 used them 1–3 times per week (weekly), 39 used them 1–3 times per month (monthly), and 7 used them less than once per month (rarely). The low use of LLMs for exam preparation may indicate students' unfamiliarity with their application in this context, or a general perception that LLMs are not suited for this task. For example, the monthly frequency corresponds with the timing of group report assignments in ICE courses.

Overall, these findings suggest that LLMs facilitate ICE self-study, especially for general tasks such as homework assistance, while their application in specialised areas such as simulation coding and data analysis remains in development. The varying frequency of LLM usage reflects different levels of reliance on

these tools in ICE self-study. Enhancing teacher guidance and promoting LLMs more widely may further increase their usage and expand their applications.

3.1.4 | Post-Usage Feedback About LLMs in ICE Self-Study Among Respondents

Dimension 4 (Q15–Q17) assessed post-usage feedback on LLMs in ICE self-study, including advantages and challenges, as detailed in Table 4. Regarding advantages (Q15), respondents reported that LLMs boosted homework and project productivity (73.4%), aided comprehension of complex topics (57.1%) and facilitated rapid reference material access (33.7%). Only 4% of respondents considered LLMs supported better preparation for exams, consistent with finding in Dimension 3 regarding their less usage in exam preparation. Additionally, 17.3% of respondents, primarily fourth-year undergraduates, believed LLMs improved technical writing, particularly for thesis preparation.

However, respondents faced several challenges with LLMs. 50.0% of respondents believed the models sometimes provided inaccurate information, and 34.7% noted their limited effectiveness in providing ICE-specific knowledge. This deficiency echoes concerns in the literature that current LLMs lack expertise in engineering disciplines (Sobieszek and Price 2022). Academic integrity is another significant issue. Approximately 88% of respondents who found LLMs helpful for academic

TABLE 4 | Advantages and limitations of LLMs usage in ICE self-study.

Question	Options	Number of votes	Percentage
What advantages have you observed from using LLMs in ICE-related coursework? (Q15: Multiple-choice question)	Increased productivity in homework & projects	72	73.4%
	Enhanced understanding of complex topics	56	57.1%
	Rapid reference materials access	33	33.7%
	Better preparation for exams	4	4.1%
	Improved technical writing	17	17.3%
	None of these	0	0.0%
	Other (please specify)	0	0.0%
What challenges or limitations have you encountered with LLMs in ICE-related coursework? (Q16: Multiple-choice question)	Complexity of use	54	55.1%
	Inaccurate information	49	50.0%
	Limited ICE-specific knowledge	34	34.7%
	Data privacy concerns	18	18.4%
	Access costs	46	46.9%
	Integration challenges with ICE tools	21	24.4%
	Academic integrity concerns	15	15.3%
	None of these	0	0.0%
Do you believe LLMs are suitable for academic studies in ICE? (Q17: Single-choice question)	Yes, they are suitable as supplementary tools	78	79.6%
	Not sure/It depends on the sub-discipline	1	1.0%
	No, they are not suitable	19	19.4%

writing (15 students from Q16) expressed concerns about academic integrity (17 students from Q15). This aligns with concerns in recent literature (Mhlana 2023). Furthermore, challenges included operational difficulties, with 18.4% of respondents concerned about data privacy, 55.1% citing the complexity of use, and 46.9% highlighting high access costs. Regarding the suitability of LLMs for ICE self-study, 79.6% of respondents considered them valuable adjuncts to ICE courses, while 19.4% were uncertain, and 1.0% deemed them unsuitable. Interviews with students who answered 'Not sure' or 'No' in Q17 expressed concerns that the limited capabilities in image/video analysis and Chinese processing of LLMs may reduce their effectiveness in analysing control systems.

Based on this survey, several insights emerge. LLMs demonstrate significant potential in improving learning efficiency and providing abundant resources to meet personalised learning needs. However, current LLMs still face challenges in addressing the demands of professional education. To better support deeper learning in ICE courses, teachers should gradually explore appropriate application scenarios and effectively integrate LLMs into teaching. Furthermore, developers should prioritise building specialised domain knowledge bases and strengthening data privacy protections to maximise their value in engineering education.

3.2 | Performance Assessment of GPT Models on ICE Simulated Exam

The questionnaire survey reveals that ChatGPT is the most frequently used model among ICE students. The study subsequently assessed its capacity to address ICE-specific problems through a simulated ICE exam. The exam comprised 342 standardised questions with reference answers, derived from multiple ICE courses on the Chinese MOOC platform. The performance of GPT-3.5 and GPT-4 models was compared across these questions to assess their accuracy and expertise in the ICE domain.

As shown in Table 5, GPT-3.5 correctly answered 234 out of 342 questions, achieving an accuracy of 68.42%, whereas GPT-4 achieved a higher accuracy of 82.16% by correctly answering 281 questions. The results were obtained without model fine-tuning and only with Chinese prompts, indicating GPT-4's enhanced ability to handle complex ICE-related questions. Further analysis in Table 6 shows that GPT-4 correctly answered 214 of the 234 questions that GPT-3.5 also answered correctly, and improved on 67 of the 108 questions that GPT-3.5 answered incorrectly. However, 41 questions were not answered correctly by either model. These findings suggest that, although GPT-4 shows notable improvements, it still has limitations in addressing certain question types.

TABLE 5 | Performance comparison of GPT-3.5 and GPT-4 on ICE simulated exam.

Variable	GPT-3.5	GPT-4
Correct answer	234	281
Incorrect answer	108	61
Accuracy	68.42%	82.16%

To more accurately evaluate model expertise in the simulated ICE exam, the ICE questions were categorised into specific sub-specialties. These sub-specialties are defined as follows:

- *Fundamental Concepts*: Core principles include feedback, stability, controllability, observability and the dynamic behaviour of systems.
- *Mathematical Modelling*: The process of representing physical systems using differential equations, transfer functions or block diagrams to characterise their dynamic behaviour. These models often include linear, nonlinear, time-invariant and time-varying system representations.
- *Measurement Instruments*: Devices used to acquire accurate data from systems, such as sensors for position, velocity, temperature, pressure and flow. High-precision instrumentation ensures reliable feedback for control systems.
- *Controllers and Control Laws*: The design and implementation of control laws determine how systems respond to disturbances and achieve desired performance. Includes proportional-integral-derivative (PID) controllers, proportional-integral (PI) controllers, proportional-derivative (PD) controllers and model predictive control (MPC).
- *Simple Control Systems*: Systems involve single-input, single-output (SISO) systems, where proportional, integral or derivative control ensures steady-state performance and basic stability.
- *Complex Control Systems*: Systems involve multi-input, multi-output (MIMO) systems with interactions among variables. Examples include cascade control system, ratio control system, feedforward-feedback control system, uniform control system, selective control system, split-range control system and dual control system.
- *Advanced Control Systems*: Systems involve modern techniques like adaptive control, optimal control and nonlinear control to manage uncertainty, time-varying parameters or nonlinear dynamics.
- *Classical Control Applications*: These applications rely on traditional control techniques. Examples include thermostats, level control systems, distillation towers.

The ICE questions were categorised based on these sub-specialties, with the percentages given for each category

TABLE 6 | Accuracy comparison of GPT-3.5 and GPT-4 in ICE simulated exam (2×2 contingency table).

Variable	GPT-4 correct answer	GPT-4 incorrect answer	Total
GPT-3.5 correct answer	214	20	234
GPT-3.5 incorrect answer	67	41	108

(Figure 1, listed from highest to lowest): Simple Control Systems (25.15%), Complex Control Systems (18.13%), Fundamental Concepts (17.84%), Mathematical Modelling (11.70%), Measurement Instruments (9.06%), Controllers and Control Laws (7.60%), Advanced Control Systems (5.85%) and Classical Control Applications (4.68%).

The comparative analysis of GPT-3.5 and GPT-4 models reveals significant performance differences across various ICE sub-specialties. As illustrated in Figure 2, the accuracy of GPT-3.5 across sub-specialties is as follows: Fundamental Concepts (78.69%), Mathematical Modelling (77.50%), Measurement Instruments (54.84%), Controllers and Control Laws (65.38%), Simple Control Systems (73.26%), Complex Control Systems (64.52%), Advanced Control Systems (60.00%), and Classical Control Applications (37.50%). In comparison, GPT-4 demonstrates marked improvements in all sub-specialties, achieving the following accuracies: Fundamental Concepts (86.89%), Mathematical Modelling (90.00%), Measurement Instruments (67.74%), Controllers and Control Laws (84.62%), Simple Control Systems (87.21%), Complex Control Systems (75.81%), Advanced Control Systems (75.00%), and Classical Control Applications (75.00%). The accuracy improvement of GPT-4 over GPT-3.5 ranges from 8.2% to 37.5%. Additionally, GPT-4 exhibits greater stability, with accuracy fluctuations limited to a range of 22.26% (from 67.74% to 90.00%), compared to GPT-3.5's broader fluctuation of 41.19% (from 37.50% to 78.69%). This indicates that GPT-4 not only performs more accurately but also maintains greater consistency in solving ICE-related problems across diverse sub-specialties.

A detailed analysis of question type performance highlights the advancements made by GPT-4. The ICE questions are categorised into four types, with their respective proportions shown in Figure 3a: Multiple Choice (38.60%), True/False (27.49%), Fill-in-the-Blank (17.84%), and Multiple Response (16.08%). The accuracy of GPT-3.5 for these question types is as follows: Multiple Choice (71.97%), True/False (73.40%), Fill-in-the-Blank (72.13%) and Multiple Response (47.27%). In contrast, GPT-4 achieves significantly higher accuracy: Multiple Choice (85.61%), True/False (85.11%), Fill-in-the-Blank (85.25%) and Multiple Response (65.45%). The largest improvement is observed in

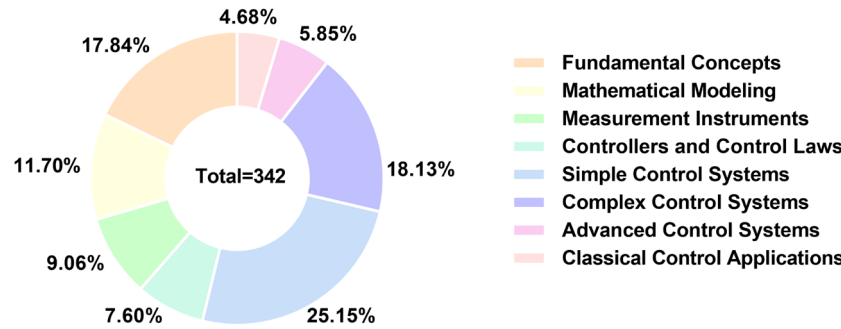


FIGURE 1 | Distribution of the 342-question dataset by ICE sub-specialties: Fundamental Concepts, Mathematical Modelling, Measurement Instrumentations, Controllers and Control Laws, Simple Control Systems, Complex Control Systems, Advanced Control Systems and Classical Control Applications.

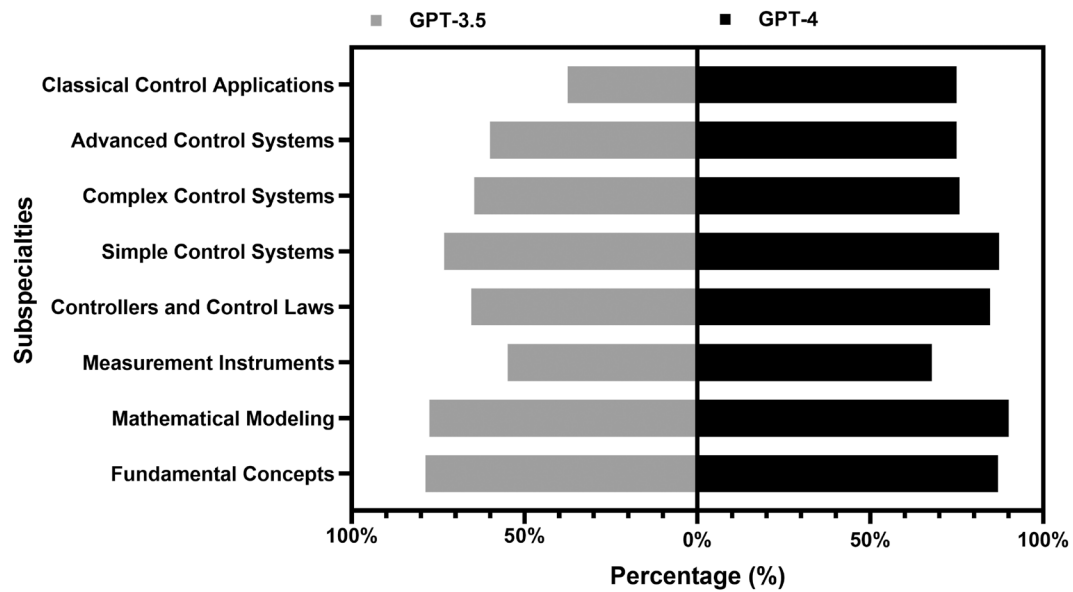


FIGURE 2 | Accuracy comparison between GPT-4 and GPT-3.5 across various ICE sub-specialties.

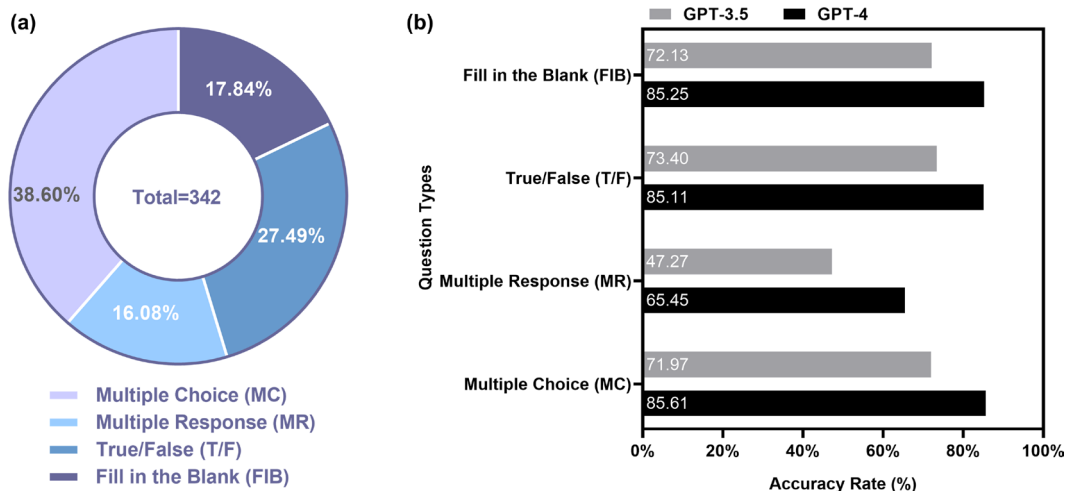


FIGURE 3 | (a) Distribution of the 342-question dataset by question type: Multiple Choice (MC), Multiple Response (MR), True/False(T/F), and Fill-in-the-Blank (FIB). (b) Accuracy comparison of GPT-4 and GPT-3.5 across different question types in the ICE simulated exam.

Multiple Response questions, with an 18.18% increase. Other question types also show notable increases: 13.64% for Multiple Choice, 13.12% for Fill-in-the-Blank, and 11.71% for True/False questions.

The comparison of GPT-3.5 and GPT-4 in a simulated ICE exam reveals notable differences in three dimensions: overall accuracy, sub-specialty accuracy, and question type accuracy. These disparities can likely be attributed to variations in training datasets and model architectures. The results emphasise the importance of considering such factors when selecting and applying LLMs into educational settings to optimise their effectiveness.

4 | Discussion

4.1 | Influences on LLMs Application in ICE Courses

Based on results from the questionnaire and interviews, ICE undergraduates generally demonstrate a high awareness of LLMs, especially ChatGPT. However, notable differences were observed in popularity of different models, likely influenced by marketing efforts and personal usage habits. LLMs are commonly used for homework and problem-solving but are less adopted for complex tasks like simulation coding and data analysis. This difference may be due to personalised learning needs, task complexity and familiarity with LLMs. Despite the widespread adoption of LLMs, challenges persist. While 79.6% of respondents expressed a positive attitude toward using LLMs for ICE self-study, some noted deficiencies in the subject-specific expertise of these models. These limitations occasionally result in inaccurate responses, particularly when addressing specialised ICE topics. For example, in a simulated ICE exam, accuracy, ChatGPT demonstrated good accuracy, with GPT-3.5 achieving 68% and GPT-4 achieving 82%. However, both models struggled with certain questions, highlighting the need for further improvement, including model fine-tuning or guidance from teachers. Additionally, students expressed concerns about several factors that impact

the adoption of LLMs. Key issues include data privacy, usage costs, and academic integrity.

Based on these results, four key factors influence the LLMs application in ICE courses:

1. *Marketing Influence*: The popularity of LLMs is significantly affected by marketing efforts. Stronger promotion enhances awareness and adoption of LLMs among students. For example, Chinese models like Baidu's Ernie Bot, Alibaba's Tongyi and iFLYTEK's Spark enjoy high popularity among Chinese students due to effective localization and widespread promotion on Chinese social platforms.
2. *Technological Proficiency*: The ability to effectively use LLMs directly impacts learning outcomes. Students with limited skills often face difficulties in formulating precise queries or assessing response accuracy.
3. *Personal Learning Needs*: Personal learning needs range from basic tasks (like homework assistance) to advanced tasks (like simulation coding and data analysis). The effectiveness of LLMs often depends on the complexity of the task and the student ability to interact with the models.
4. *Trust in LLMs*: The level of trust that students place in LLMs directly influences their willingness to follow the responses provided by these models. A lack of trust can reduce engagement with the tool, while overreliance without critical evaluation risks the acceptance of inaccurate or misleading outputs.

4.2 | Strategies for Optimising LLMs Application in ICE Courses

Most respondents wanted to use LLMs in ICE learning more effectively through educators-led approaches. To achieve this goal, educators can make improvements in the following four aspects:

1. Establishing an LLM platform that offers comprehensive resources for self-study using LLMs. It should include detailed tutorials on various LLMs and knowledge bases for ICE courses. This approach enables student to become more familiar with LLMs and develop their own model training.
2. Providing technical training to equip students with essential skills required for effectively using LLMs in their self-study process. It should cover topics such as basic interface operation, effective communicating and training commands. This approach help student in enhancing their technological proficiency when using LLMs for ICE self-study.
3. Conducting regular assessments on the learning progress in using LLMs, paying attention to any difficulties they may encountered when using these models, and provide personalised guidance and advice according. This approach can help students to learn how to use LLMs to meet different personal learning needs.
4. Establishing a trust mechanism by guiding students in understanding the limitations of LLMs, and helping them evaluate the reliability of answers generated by these models independently. Meanwhile, educators should encourage students to ask questions autonomously and share insights. This approach will build trust in the use of LLM tools.

Through these improvements, educators can help students to make better use of LLMs for ICE self-study, as well as enhance their technical application skills and professional knowledge understanding.

4.3 | Future Application of LLMs in ICE Courses

ChatGPT is just one example of LLMs. Others, like Baidu's Ernie Bot, Alibaba's Tongyi, iFLYTEK's Spark, Anthropic's Claude, and Meta's LLaMA, have shown promise in education. These models offer user-friendly application programming interfaces for open innovation and customization. As LLMs continue to evolve, they will be trained on more varied datasets, enabling them to provide specialised knowledge, such as in ICE. Some models can also process visual information, like images and videos. This enables them to analyse visual data like control system diagrams, process response curves, signal flow graphs, as well as related videos, thereby expanding their applicability in the field of ICE.

4.4 | Limitations of This Study

The limitations of this study primarily lie in the restricted survey respondents and the limited question bank. Firstly, the survey respondents were exclusively undergraduates majoring in ICE at the University of Shanghai for Science and Technology. This fact restricts the generalizability of the survey results. Secondly, the questions used in the simulated ICE exam were sourced from China MOOC platforms, and excluded questions containing visual elements. Therefore, this simulated exam may not comprehensively represent the diversity of actual qualification exams,

making it difficult to conclusively determine whether ChatGPT would meet the qualifications for instrument engineers and control engineers. However, despite these limitations, this study remains significant as it provides valuable insights into the awareness and usage habits of students toward LLMs in ICE self-study, as well as demonstrates the ability of LLMs to analyse ICE-related information.

5 | Conclusion

The integration of LLMs in education is still in its early stages. This paper examines ICE as a case study. A survey was conducted to evaluate the awareness and usage habits of ICE undergraduates at the University of Shanghai for Science and Technology, regarding LLMs in self-study. Results showed that ChatGPT is the most popular model, used by 72.4% of respondents. Furthermore, 79.6% of ICE students had a positive attitude toward LLMs in ICE self-study. To further assess the effectiveness of ChatGPT in ICE self-study, we tested GPT-3.5 and GPT-4 models in a simulated ICE exam. GPT-4 showed better understanding of complex control information with an 82.16% accuracy, surpassing GPT-3.5's 68.42%. GPT-3.5's accuracy varied by 41.19% across ICE sub-specialties, while GPT-4's variation was only 22.26%, indicating more consistent performance across sub-specialties. For different question types, GPT-3.5's accuracy varied between 47.27% and 73.40%, whereas GPT-4's ranged from 65.45% to 85.61%, highlighting notable performance differences. Finally, we identified key factors influencing students' use of LLMs for ICE self-study, such as marketing efforts, technological proficiency, personalised learning needs, and trust in LLMs. Additionally, we proposed suggestions to educators for improving LLMs' effectiveness in ICE learning, including establish an LLM platform, provide technical training, conduct regular assessments, and establish a trust mechanism.

Acknowledgements

This study was supported by Shanghai Municipal Education Commission Key Course Construction Project.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Abas, M. A., S. E. Arumugam, M. M. Yunus, and K. R. M. Rafiq. 2023. "ChatGPT and Personalized Learning: Opportunities and Challenges in Higher Education." *International Journal of Academic Research in Business and Social Sciences* 13, no. 12: 3936–3945. <https://doi.org/10.6007/IJARBS/v13-i12/20240>.
- Adetayo, A. J., M. O. Aborisade, and B. A. Sanni. 2024. "Microsoft Copilot and Anthropic Claude AI in Education and Library Service." *Library Hi Tech News*. <https://doi.org/10.1108/LHTN-01-2024-0002>.
- Alhazmi, S., S. Khan, and M. H. Syed. 2023. "Learning-Related Sentiment Detection, Classification, and Application for a Quality

- Education Using Artificial Intelligence Techniques." *Intelligent Automation & Soft Computing* 36, no. 3: 3487–3499. <https://doi.org/10.32604/iasc.2023.036297>.
- Ali, R., O. Y. Tang, I. D. Connolly, et al. 2023. "Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank." *Neurosurgery* 93, no. 5: 1090–1098. <https://doi.org/10.1227/neu.0000000000002551>.
- Angel, M. C., J. B. Rinehart, M. P. Cannesson, and P. J. A. Baldi. 2022. "Clinical Knowledge and Reasoning Abilities of AI Large Language Models in Anesthesiology: A Comparative Study on the American Board of Anesthesiology Examination." *medRxiv* 139, no. 2: 349–356.
- Ansar, W., S. Goswami, and A. Chakrabarti. 2024. "A Survey on Transformers in NLP With Focus on Efficiency." *arXiv*. <https://doi.org/10.48550/arXiv.2406.16893>.
- Chalkidis, I. 2023. "Chatgpt May Pass the Bar Exam Soon, but Has a Long Way to Go for the Lexglue Benchmark." *arXiv*. <https://doi.org/10.48550/arXiv.2304.12202>.
- Chen, J., Z. Liu, X. Huang, et al. 2024. "When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities." *arXiv*. <https://doi.org/10.48550/arXiv.2307.16376>.
- de Winter, J. C. F. 2023. "Can ChatGPT Pass High School Exams on English Language Comprehension?" *International Journal of Artificial Intelligence in Education* 34: 915–930. <https://doi.org/10.1007/s40593-023-00372-z>.
- Finnie-Ansley, J., P. Denny, B. A. Becker, A. Luxton-Reilly, and J. Prather. 2022. "The Robots Are Coming: Exploring the Implications of Openai Codex on Introductory Programming." In *Proceedings of the 24th Australasian Computing Education Conference*. Virtual Event, Australia: Association for Computing Machinery.
- Fitria, T. N. 2023. "Artificial Intelligence (AI) Technology in OpenAI ChatGPT Application: A Review of ChatGPT in Writing English Essay." *ELT Forum: Journal of English Language Teaching* 12: 44–58.
- Freitas, P. M., and L. M. Gomes. 2023. *Does ChatGPT Pass the Brazilian Bar Exam? Progress in Artificial Intelligence*. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49011-8_11.
- Ghavidel, H. A., A. Zouaq, and M. C. Desmarais. 2020. "Using BERT and XLNET for the Automatic Short Answer Grading Task." *International Conference on Computer Supported Education* 1: 58–67. <https://doi.org/10.5220/0009422400580067>.
- Goyal, T., J. J. Li, and G. Durrett. 2022. "News Summarization and Evaluation in the Era of GPT-3." *arXiv* 2, no. 4. <https://doi.org/10.48550/arXiv.2209.12356>.
- Haleem, A., M. Javaid, and R. P. Singh. 2022. "An Era of ChatGPT as a Significant Futuristic Support Tool: A Study on Features, Abilities, and Challenges." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 2: 100089.
- Hendy, A., M. Abdelrehim, A. Sharaf, et al. 2023. "How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation." *arXiv*. <https://doi.org/10.48550/arXiv.2302.09210>.
- Imran, M., and N. Almusharraf. 2023. "Analyzing the Role of ChatGPT as a Writing Assistant at Higher Education Level: A Systematic Review of the Literature." *Contemporary Educational Technology* 15, no. 4: ep464. <https://doi.org/10.30935/cedtech/13605>.
- Jacob Devlin, M.-W. C., K. Lee, and K. Toutanova. 2019. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 4171–4186. Minneapolis, Minnesota publisher: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>.
- Javaid, M., A. Haleem, R. P. Singh, S. Khan, and I. H. Khan. 2023. "Unlocking the Opportunities Through ChatGPT Tool Towards Ameliorating the Education System." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3, no. 2: 100115. <https://doi.org/10.1016/j.t.bench.2023.100115>.
- Jeon, J., and S. Lee. 2023. "Large Language Models in Education: A Focus on the Complementary Relationship Between Human Teachers and ChatGPT." *Education and Information Technologies* 28, no. 12: 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>.
- Kale, M., and A. Rastogi. 2020. "Text-To-Text Pre-Training for Data-To-Text Tasks." *arXiv*. <https://doi.org/10.48550/arXiv.2005.10433>.
- Karanikolas, N., E. Manga, N. Samaridi, E. Tousidou, and M. Vassilakopoulos. 2024. *Large Language Models versus Natural Language Understanding and Generation*, 278–290. Lamia, Greece: In Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics. Association for Computing Machinery.
- Katz, D. M., M. J. Bommarito, S. Gao, and P. Arredondo. 2024. "GPT-4 Passes the Bar Exam." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 382, no. 2270: 20230254. <https://doi.org/10.1098/rsta.2023.0254>.
- Kheddar, H. 2024. "Transformers and Large Language Models for Efficient Intrusion Detection Systems: A Comprehensive Survey." *arXiv*. <https://doi.org/10.48550/arXiv.2408.07583>.
- Kosar, T., D. Ostojić, Y. D. Liu, and M. Mernik. 2024. "Computer Science Education in ChatGPT Era: Experiences From an Experiment in a Programming Course for Novice Programmers." *Mathematics* 12, no. 5: 629.
- Li, S., R. Carlson, and C. J. Potts. 2022. "Systematicity in GPT-3's Interpretation of Novel English Noun Compounds." *arXiv*. <https://doi.org/10.48550/arXiv.2210.09492>.
- Limo, F. A. F., D. R. H. Tiza, M. M. Roque, et al. 2023. "Personalized Tutoring: ChatGPT as a Virtual Tutor for Personalized Learning Experiences." *Przestrzen Społeczna* 23, no. 1: 293–312.
- Mhlanga, D. 2023. "Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning." In *Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning, in FinTech and Artificial Intelligence for Sustainable Development: The Role of Smart Technologies in Achieving Development Goals*, 387–409. Cham: Springer.
- Moore, S., H. A. Nguyen, N. Bier, T. Domadia, and J. Stamper. 2022. "Assessing the Quality of Student-Generated Short Answer Questions Using GPT-3." In *European Conference on Technology Enhanced Learning*. Cham: Springer International Publishing.
- Newton, P., and M. Xiromeriti. 2024. "ChatGPT Performance on Multiple Choice Question Examinations in Higher Education. A Pragmatic Scoping Review." *Assessment & Evaluation in Higher Education* 49, no. 6: 781–798. <https://doi.org/10.1080/02602938.2023.2299059>.
- Sadeq, M. A., R. M. F. Ghorab, M. H. Ashry, et al. 2024. "AI Chatbots Show Promise but Limitations on UK Medical Exam Questions: A Comparative Performance Study." *Scientific Reports* 14, no. 1: 18859. <https://doi.org/10.1038/s41598-024-68996-2>.
- Shahzad, M. F., S. Xu, and I. Javed. 2024. "ChatGPT Awareness, Acceptance, and Adoption in Higher Education: The Role of Trust as a Cornerstone." *International Journal of Educational Technology in Higher Education* 21, no. 1: 46. <https://doi.org/10.1186/s41239-024-00478-x>.
- Sobieszek, A., and T. J. M. Price. 2022. "Playing Games With AIs: The Limits of GPT-3 and Similar Large Language Models." *Minds and Machines* 32, no. 2: 341–364.
- Somers, R., S. Cunningham-Nelson, and W. Boles. 2021. "Applying Natural Language Processing to Automatically Assess Student Conceptual Understanding From Textual Responses." *Australasian Journal of Educational Technology* 37, no. 5: 98–115. <https://doi.org/10.14742/ajet.7121>.

- Sun, Y., S. Wang, S. Feng, et al. 2021. "ERNIE 3.0: Large-Scale Knowledge Enhanced Pre-Training for Language Understanding and Generation." *arXiv*. <https://doi.org/10.48550/arXiv.2107.02137>.
- Sun, Y., S. Wang, Y. Li, et al. 2019. "ERNIE: Enhanced Representation Through Knowledge Integration." *arXiv*. <https://doi.org/10.48550/arXiv.1904.09223>.
- Sun, Y., S. Wang, Y. Li, et al. 2020. "Ernie 2.0: A Continual Pre-Training Framework for Language Understanding." *arXiv*. <https://doi.org/10.48550/arXiv.1907.12412>.
- Vaswani, A., N. Shazeer, N. Parmar, et al. 2017. "Attention Is All You Need." *arXiv*. <https://doi.org/10.48550/arXiv.1706.03762>.
- Wang, S., T. Xu, H. Li, et al. 2024. Large Language Models for Education: A Survey and Outlook.
- Xu, W., and F. Ouyang. 2022. "The Application of AI Technologies in STEM Education: A Systematic Review From 2011 to 2021." *International Journal of STEM Education* 9, no. 1: 59. <https://doi.org/10.1186/s40594-022-00377-5>.
- Zhang, R. 2022. "Personalized Course Recommendation Method Based on Learner Interest Mining in Educational Big Data Environment." *Scientific Programming* 2022, no. 1: 9943965. <https://doi.org/10.1155/2022/9943965>.
- Zhilin Yang, Z. D., Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le. 2019. "XLNet: Generalized Autoregressive Pretraining for Language Understanding." *arXiv*. <https://doi.org/10.48550/arXiv.1906.08237>.