



Research on the Innovation Path of Big Data Application Professional Courses Based on AIGC Big Model

Yating Zheng*, Fang Meng

GUANGZHOU CITY POLYTECHNIC, Guangzhou, China

*zhengyating@gcp.edu.cn, mengfang@gcp.edu.cn

Abstract. As an important branch of artificial intelligence, the AIGC (Artificial Intelligence Generated Content) model has brought profound impacts to various industries with its powerful data processing and content generation capabilities. How to effectively introduce and integrate the AIGC big model in vocational big data technology courses is still in the exploratory stage of application in the teaching field. This article combines the research and evaluation of the AIGC big model with the specific integration of big data technology courses to develop suitable teaching methods and strategies, such as project driven, case studies, etc.

Keywords: AIGC, big models, big data courses, code generation.

1 Introduction

In the report, it is pointed out that we should actively promote the "smart empowerment" action, accelerate the deep integration of artificial intelligence and key industries, develop a comprehensive three-dimensional teaching field and a big data based online learning education platform in the field of smart education, develop smart education assistants, establish a fast, comprehensive, and accurate education analysis system, establish a learner centered education environment, provide precise educational services, and achieve customized daily education and lifelong education.

The AIGC (Artificial Intelligence Generated Content) model, as an important branch of artificial intelligence, has brought profound impacts to various industries with its powerful data processing and content generation capabilities^[1]. The large models on the market that empower programming education products can provide common code error correction, code interpretation, question answering and other programming tutoring services for lower grade computer students with high accuracy, assisting teachers in efficient programming teaching and guiding students in independent programming and learning. In response to the complex programming problems that senior computer students may encounter, especially in the big data technology courses of vocational colleges, students need to solve the code interpretation and generation problems related to mathematical modeling and logical analysis. For example, they need to learn and master how to process and analyze large-scale datasets,

and master skills such as data processing, data mining, and data analysis in application scenario cases, which are still being actively explored. Based on the above background, this study explores and studies the paradigm of how AIGC empowers the construction of big data technology professional courses^[2].

2 Research Status

At present, many universities and research institutions in China are actively conducting research on AIGC technology in the field of education, exploring how to apply it to personalized learning, intelligent evaluation and feedback, virtual laboratories and simulation environments, etc., in order to improve teaching quality and learning effectiveness. These studies not only contribute to promoting innovation and development in the field of education, but also provide students with more personalized, efficient, and interesting learning experiences. Some educational technology companies are also actively developing education products based on AIGC technology, such as intelligent teaching systems, online learning platforms, etc.

However, there is currently relatively little research on the application of AIGC big models in the construction of big data technology courses. On the one hand, the AIGC big model, as an emerging technology, combines advanced technologies from multiple fields such as artificial intelligence, big data, and machine learning. The novelty and complexity of this technology pose certain challenges for both research and application. How to effectively introduce and integrate the AIGC big model in the construction of big data technology courses requires teachers to have interdisciplinary knowledge and skills, which increases the difficulty of research. Its application in the field of teaching is still in the exploratory stage, requiring more research and practice to verify its effectiveness and feasibility. On the other hand, due to the relatively new application of the AIGC large model, there is a lack of relevant teaching resources and cases. This makes it difficult to find suitable teaching cases and experimental data to support the teaching of AIGC big models in the construction of big data technology courses^[6]. The AIGC big model involves regulatory and ethical issues such as data privacy and intellectual property in its application. How to ensure that teaching and research activities comply with relevant regulations and ethical requirements in the construction of big data technology courses is also a problem that needs to be studied.

3 Research Path

3.1 Study the AIGC Large Model

The mainstream AIGC big model in the market

ChatGPT, the latest version currently uses reinforcement learning optimization models, which are continuously trained and corrected to make larger models more powerful^[3].

LLaMA, (English: Large Language Model Meta AI) is an open-source large model released by Meta AI to promote research in the field of natural language processing

(NLP). In multiple benchmark tests, the 13 billion parameter model run by LLaMA outperformed the model provided by GPT-3 in most NLP benchmark tests. Especially, LLaMA's pre training dataset is a large-scale, diverse, and extensive dataset that contains a large amount of text, image, and video data, which comes from various online resources such as Wikipedia, social media, news websites, etc.

The main core functions of the.

Spark Model include machine translation, text summarization, syntax checking, code ability, multimodal interaction, and sentiment analysis. It can not only process multiple sources of information such as long text, long graphics, and long speech, but also provide professional and accurate answers in various industry scenarios

Baidu ERNIE Bot, is able to carry out multiple rounds of dialogue, create text content, provide knowledge Q&A, translation and other natural language processing tasks

Alibaba Tongyi Qianwen, mainly serves multiple fields such as e-commerce, finance, and education, conducting tasks such as text generation, code writing, intelligent Q&A, and inference analysis^[4].

3.2 Evaluate the AIGC Big Model

The AIGC big model is crucial for the effectiveness of big data professional courses. Different models are suitable for different scenarios and tasks, and have different advantages and limitations in understanding and generating content^[5]. Therefore, it is necessary to choose the appropriate model according to the specific needs of the course, in order to improve the adaptability and effectiveness of teaching. The specific implementation plan is as follows: Firstly, clarify the specific requirements of the big data course, including teaching content, target student groups, and the types and formats of teaching resources. Taking the course "Big Data Visualization" as an example, it is necessary to clarify the following aspects:

The target students are sophomores in big data, with a certain programming foundation and project development experience^[6]. The current main problem that needs to be solved is that teachers lack a large number of datasets that simulate real production environments, resulting in unclear visualization effects; Students lack the ability to correct long and complex code errors, as well as the ability to interpret and generate code for complex application problems involving mathematical modeling and logical analysis. Combined with the functions and characteristics, technical strength, data privacy and security, application scenarios, scalability and flexibility, compatibility, and future development trends of mainstream AIGC models or products on the market. Select an AIGC big model platform that is suitable for the big data technology course, and conduct in-depth evaluation to ensure that it can meet the teaching needs of the course^[7].

Optimizing the model. Improve the performance and accuracy of the model. This includes selecting appropriate hyperparameters, adjusting model structure, optimizing training process, etc. Evaluate the performance of the selected AIGC big model in generating big data course content through actual testing or trial. This can include evaluating the accuracy, adaptability, generation speed, and user friendliness of the

model. Based on the evaluation results, make necessary optimizations and adjustments to the AIGC large model. This includes adjusting the parameters of the model, adding training data, or improving the algorithm of the model. Continuously collect feedback from students and teachers during the teaching process, and continuously improve and optimize the AIGC model based on the feedback results. This helps to ensure that the curriculum always stays in sync with the latest big data technologies and teaching trends.

3.3 Design and Develop Big Data Professional Courses

Develop teaching resources suitable for big data professional courses, such as textbooks, courseware, datasets, online courses, etc. 2.2 Exploration of Generative Project Teaching Mode Based on AIGC Big Model Technology, such as Simulating Big Data Application Scenario Experiments, Multidimensional Student Portraits, Big Data Case Analysis, etc

Simulated Big Data Application Scenario Experiment.

Build a highly realistic big data processing and analysis environment using AIGC big model technology. This environment can simulate real big data application scenarios, including various links such as data collection, storage, processing, analysis, and visualization. Students can conduct various experimental operations, observe and analyze experimental results in this environment.

Recommended Topics for Big Data Projects.

Building a multidimensional student profile based on students' interests and abilities, the AIGC big model can recommend suitable topics for big data projects, helping students find research directions that suit them. The AIGC big model provides students with rich data resources and analysis tools, supports project research and data analysis, and can monitor project progress and student performance in real-time, providing personalized guidance and suggestions for students.

Provide a Big Data Application Case Library.

By providing a rich case library through the AIGC big model, students are provided with diverse learning materials, and representative and practical cases are selected. These cases cover various aspects of big data processing and analysis, including data collection, storage, processing, analysis, and visualization. They can help students analyze key problems and difficulties in the cases, and provide ideas and methods for solving problems. Simultaneously supporting students to share and exchange their insights and experiences on case studies online, promoting knowledge sharing and cognitive collision.

Carry out Teaching Activities According to the Established Teaching Methods and Strategies, Including Classroom Teaching, Experimental Teaching, Practical Activities, etc.

Collect feedback from students and teachers during the teaching process, evaluate teaching effectiveness, and adjust and optimize teaching plans in a timely manner

3.4 Continuous Improvement and Optimization

Continuously monitor the new developments in AIGC big model technology and big data technology to ensure that the curriculum system keeps up with the times. Collect application feedback: Collect feedback from universities or educational institutions that have adopted the new curriculum system to understand the actual application effects. Optimization and Update: Based on feedback and technological development, continuously optimize and update the curriculum system and teaching resources to improve the quality of education and the effectiveness of talent cultivation.

4 The Challenge of AIGC Big Model in the Construction of Big Data Technology Curriculum

4.1 Code Accuracy

AIGC large model technology trains through massive amounts of data, which may contain One sidedness, bias, or errors may be influenced by various factors such as training data, model architecture, and parameter settings, resulting in uneven quality of generated code. If the model does not fully learn the best practices and standards of code writing, the generated code may have potential security risks or performance issues.

4.2 Safety Issues

The code generated by the AIGC large model may contain known or unknown security vulnerabilities, which can be exploited by attackers and cause damage to the system. For example, the generated code may contain unverified inputs, unsafe function calls, or configuration errors, all of which can increase the risk of system attacks

4.3 Data Privacy Issues

For example, in the data collection stage, the user inputs sensitive information such as personal information, preferences, and behaviors. If these data are not properly protected, there is a risk of being accessed and abused by third parties; In the data processing stage, if there is bias in the training data, it may lead to biased content generated by the model, resulting in unfair treatment or discrimination against specific user groups. In the data storage stage, if the data storage system has security vulnerabilities or is attacked by hackers, it may lead to the leakage of a large amount of user privacy

data, causing serious losses to users. In the data usage stage, if the content generated by the AIGC big model is improperly used by users (such as for fraud, spreading rumors, extortion, etc.), it may have a negative impact on others or society, and may also infringe on the privacy rights of others. The AIGC big model may reference and analyze a large number of existing works when generating content, thus posing a risk of infringement on the intellectual property rights of others.

5 Conclusion

Introducing the AIGC big model into the construction of big data technology courses not only brings new research perspectives and tools to the field of educational technology, but also expands the boundaries of educational technology research. This helps to promote theoretical innovation and practical exploration in the field of educational technology, providing new ideas and directions for the development of educational technology. The application of the AIGC big model provides new ideas and methods for the construction of big data professional courses. Through in-depth research on its empowerment mechanism, application effects, and other aspects, a complete theoretical system for the construction of big data professional courses can be formed, providing reference and guidance for the construction of other related professional courses.

It is only necessary to pay attention to the issues of code accuracy, security, and data privacy that still exist in AIGC big model technology. In the application process, attention should be paid to combining manual evaluation to ensure the accuracy of test results.

References

1. Xiao Fengjia. Exploration of the Application of AIGC Technology in Software Testing. *Industrial Control Computer*, 2024.82-82.
2. SURAMEERY N M S, SHAKOR M Y. Use ChatGPT to solve programming bugs[J]. *International Journal of Information Technology & Computer Engineering*, 2023, 3(1):17-22.
3. DAVIDE CASTELVECCHI. Are ChatGPT and AlphaCode going to replace programmers?[J]. *Nature*, 2022.
4. OPENAI. Content Policy [EB/OL]. (2023 -03 -23) [2023 -07 -24]. <https://help.openai.com/en/collections/3643409-dall-e-con-tent-policy>.
5. YI Z. International student perceptions of information needs and use[J]. *The Journal of Academic Librarianship*,2007,33(6):666-673.
6. GUO Yajun.Study on the Information Needs of College Students Using AIGC——Taking ChatGPT as an Example[J]. *Information Science*.ISSN 1007-7634,CN 22-1264/G2.
7. WEIZENBAUM J. ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine[J]. *Communications of the Acm*, 1983, 9(1):36-45.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

