979-8-3315-1024-4/24/$31.00 ©2024 IEEE Natural Language Processing Algorithms for Academic Content Generation  
Alfredo Paredes-Vargas  
Information Systems Engineering Universidad Peruana de ciencias  
Aplicadas UPC  
Lima, Peru u201911812@upc.edu.pe Piero Salinas-Llorca Information Systems Engineering  
Universidad Peruana de ciencias  
Aplicadas UPC  
Lima, Peru u20201C202@upc.edu.pe José Santisteban Information Systems Engineering Universidad Peruana de Ciencias Aplicadas  
Lima, Perú PCSILSAN@upc.edu.pe Abstract— This study has conducted several activities to analyze and compare the performance of five Natural Language Processing (NLP) models: GPT-3, T5, ERNIE, BERT, and XLNet for generating academic content in universitie s located in Lima, Peru. Some NLP models, including GPT and BERT, were assessed based on accuracy, quality, speed, an d customization capabilities. The research examined th e use of various algorithms for academic tasks like essay an d summary creation. The findings show that all models perform adequately; however, some excel in generating coherent content tailored to different academic styles, thereby optimizing the t ime of both students and teachers. In conclusion, the ERNIE model emerges as one of the best, as it enables the generation of diverse and high-quality content efficiently. Keywords—algorithm, SCORM, NLP, academic content, higher education. I. INTRODUCTION  
The creation of educational content encounters vari ous challenges [1]. There is an increasing demand for u pdated and adaptable materials that address the rapid changes in knowledge and technology [2]. Universities must cre ate content that is accessible, relevant, and promotes student engagement [3]. However, developing this content de mands significant time, skilled personnel, and investment in technology and authoring tools. These requirements can hinder institutions’ ability to adapt to emerging e ducational needs. The increasing demand for online learning an d the use of Learning Management Systems (LMS) necessitate co ntent that is both high-quality and compliant with intern ational standards like the Sharable Content Object Referenc e Model (SCORM). This compliance promotes the adoption of g lobal standards in content creation, fostering a more inc lusive education that meets the needs of today’s digital environment. Additionally, it allows for the reuse and interoperability of educational materials [4]. Generative Artificial Intelligence (GAI) has transf ormed academic content generation by automatically creati ng texts, images, simulations, and other educational resource s [5]. By utilizing advanced models like Generative Pre-train ed Transformer (GPT), GAI can create high-quality, ada ptable educational materials, greatly decreasing the time and costs involved in content development [6]. Additionally, it enables educators to focus on personalizing and continuousl y improving learning, providing more dynamic and inte ractive experiences for students. The study aims to evaluate and compare various algorithms to identify which ones provide greater e fficiency and quality in developing educational materials. Th is analysis aims to find suitable solutions for enhancing acade mic productivity, reducing work hours, and optimizing c oherence and accuracy in text generation. The remainder of the paper is organized as follows:  
Section 2 discusses methodology, and Section 3 prov ides the results and analysis. Finally, in Section 4, the co nclusion of the presented work is provided. II. METHODOLOGY To achieve our goal in this research, we used a dat aset specifically for this comparison and pre-trained th e models with relevant information to assist in generating a cademic content. The overall framework of the research meth od is shown in Figure 1.

Fig. 1. Steps of the research process A. Dataset Stratified sampling was used to ensure thematic div ersity and representativeness of the corpus, considering t he distribution of academic disciplines in repositorie s such as arXiv and Google Scholar. This dataset contains a w ide range of academic texts, including abstracts and full art icles across various disciplines such as engineering, social sci ences, and humanities. These resources were chosen for their r elevance in producing high-quality academic content in both Spanish and English. Having a broader range of languages fa cilitates the examination of how algorithms interact with lin guistic 2024 IEEE 4th International Conference on Advanced Learning Technologies on Education & Research (ICALTER) | 979-8-3315-1024-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICALTER65499.2024.10819214 Authorized licensed use limited to: Universidad Peruana de Ciencias Aplicadas (UPC). Downloaded on September 11,2025 at 04:08:42 UTC from IEEE Xplore. Restrictions apply. and thematic structures, which are crucial for the adaptability of generated content [7]. The selection of sources is focused on the accessib ility of academic articles that offer representative example s of conducted research. This approach ensures that the NLP algorithms are evaluated not only on the quality of the content they generate but also on their ability to adapt to various academic contexts. In addition, external validation of the results is planned through collaboration with resea rchers who replicate the study with similar data sets. Accord ing to Yin W. et al. [8], datasets that combine content from v arious sources and fields of study offer a more comprehens ive view of the models’ ability to generalize. This ability is crucial for this type of analysis. Below in Table 1, we outline the algorithms that will be compared, using metrics suc h as accuracy, linguistic quality, and content generatio n time measured in milliseconds “ms”. TABLE I. DATASET ON THE ALGORITHMS AND THEIR MULTI PLE PERFORMANCE VARIABLES Algorithm Input (Prompt) Coherence Precision Linguistic Quality Generation Time (s) Length (Words) Academic Discipline Expert Evaluation GPT-3 “Explain the theory of…” 4.5 4.8 4.7 3.5 500 Social Sciences 4.6 BERT “Describe the process…” 3.8 4.0 3.9 4.0 450 Computer Science 4.1 T5 “List the factors…” 4.0 4.2 4.1 3.8 480 Social Sciences 4.3 XLNet “Discuss the importance…” 4.2 4.5 4.3 3.7 490 Social Sciences 4.4 ERNIE “Analyze the evolution…” 4.1 4.3 4.0 4.2 470 Computer Science 4.2 When creating academic content, it is crucial to fo cus on coherence, linguistic quality, and generation time. These factors ensure that the produced material is releva nt, accurate, and easy to understand. Coherence allows for a logi cal and smooth development of ideas, while accuracy ensures that the information is correct and well-supported. A scorin g system ranging from 1 to 5 was chosen to offer more rating options for comparing the activities conducted by each algo rithm. B. Preprocessing Data preprocessing included cleaning of special characters and labels, text normalization, tokeniza tion, and stopword removal. Model selection focused on cuttin g-edge NLP technologies such as GPT-3, BERT, XLNet, T5, an d ERNIE, excluding simpler algorithms (n-grams, Marko v Chains) and less effective approaches for academic text (LSTM, RNN). Recent models such as LaMDA were not considered due to access restrictions. GPT-3, devel oped by OpenAI, is notable for its ability to produce coher ent content across various contexts. This capability stems from its broad recognition and extensive feedback, which facilitat e improvements and optimization in the way informatio n is presented [9]. BERT, developed by Google, emphasize s understanding word context bidirectionally, making it especially useful for interpretation tasks like tex t classification [10]. T5 transforms any activity or task into a text-base d problem, facilitating tasks such as summary, transl ation, or generation. XLNet, which was also developed by Goog le, employs a pretraining method that effectively captu res long- term relationships between words [11]. Finally, we have ERNIE from Baidu, which focuses on integrating stru ctured data to enhance accuracy in content generation [12] . III. RESULTS AND ANALYSIS In this section, we analyzed the performance of the models in generating academic content. The metrics evaluat ed include Precision, Recall, F1 Score, and Processing Time (m s). A comparative graph was created to analyze each metri c.  A. Precision Precision is a metric that evaluates the prediction s of an NLP model, measuring the ratio of correct predictio ns to the total predictions made [13]. In academic content ge neration, precision is essential because a highly precise mod el reduces the likelihood of producing irrelevant or incorrect  
information. This aligns with the findings in [16], which suggest that precision is particularly important wh en the goal is to minimize false positives, thereby resulting i n fewer incorrect predictions.

Fig. 2. Comparative analysis of NLP algorithms rega rding precision Figure 2 shows that GPT-3 and ERNIE exhibit outstan ding precision levels, differing from the other models b y only a Authorized licensed use limited to: Universidad Peruana de Ciencias Aplicadas (UPC). Downloaded on September 11,2025 at 04:08:42 UTC from IEEE Xplore. Restrictions apply. few decimal points. This reflects their ability to generate highly relevant academic content while minimizing t he production of unrelated information. These results align with the characteristics of GPT-3, which effectively han dles broad contexts, and ERNIE, which excels at integrating st ructured knowledge to enhance content relevance. B. Recall Sensitivity indicates a model’s capacity to correct ly identify all relevant instances within a dataset [1 4]. Liu et al.  [9] emphasizes that sensitivity is essential in NLP tasks aimed at capturing as many relevant instances as possible , particularly in generating comprehensive academic c ontent. A high level of sensitivity indicates that the mode l or algorithm effectively identifies the most relevant content.

Fig. 3. Comparative analysis of NLP algorithms rega rding sensitivity. According to Figure 3, BERT and ERNIE demonstrate a  
higher recall compared to the other models. This in dicates their superior ability to identify relevant instanc es within the evaluated data, a critical factor for generating co mprehensive and representative academic content. This recall ad vantage may be linked to BERT’s bidirectional architecture and ERNIE’s contextual capabilities. C. F1 Score It is a metric that integrates precision and sensit ivity into a harmonic measure, achieving a balance between the two [16]. Devlin et al. [10] indicate that the F1 Score is especially valuable when a trade-off is necessary to minimize errors or avoid potential false positives by omitting relevan t instances. In content generation, a high F1 Score indicates th at the model produces not only accurate text but also comp lete and relevant content.  
Fig. 4. Comparative analysis of NLP algorithms rega rding the F1 Score. Figure 4 illustrates that ERNIE outperforms the oth er models in F1-Score, establishing itself as the most balanc ed in terms of precision and recall. This strong performance re flects its ability to generate academic content that combines accuracy with adequate coverage, minimizing errors while pro ducing complete information. D. Processing Time Radford et al. [11] indicate that processing time i s a crucial factor in generating academic content, as it affect s the model’s viability for real-time or large-scale applications .

Fig 5. Comparative analysis of NLP algorithms regar ding processing time. Figure 5 highlights that while GPT-3 achieves the h ighest precision, it also has the longest processing time, which could limit its applicability for tasks requiring real-ti me results. Conversely, BERT and XLNet achieve significantly sh orter processing times, making them ideal for application s where speed is critical, without compromising acceptable performance in other metrics. Authorized licensed use limited to: Universidad Peruana de Ciencias Aplicadas (UPC). Downloaded on September 11,2025 at 04:08:42 UTC from IEEE Xplore. Restrictions apply. E. Comparative analysis of NLP algorithms

Fig. 6. Comparative analysis of NLP algorithms Figure 6 illustrates that GPT-3 achieved higher pre cision, while ERNIE distinguished itself with balanced perf ormance across all metrics, featuring a high F1 score and r elatively efficient processing times. Although more complex m odels like GPT-3 and ERNIE exhibit longer processing time s, BERT and XLNet provide a good balance between performance and efficiency. IV. CONCLUSION  
This study was conducted to evaluate and compare fi ve natural language processing (NLP) algorithms like G PT-3, BERT, T5, XLNet, and ERNIE in the generation of aca demic content, a critical task to address the needs for q uality and adaptability in higher education. These models were analyzed in terms of precision, recall, F1-Score, processing time, and fundamental metrics to ensure that the generated te xts are coherent, relevant, and effective. The results revealed that GPT-3 excelled in terms o f high precision and linguistic quality, making it ideal f or tasks prioritizing the fidelity of the generated content. BERT and XLNet demonstrated a balance between efficiency and  
processing time, making them suitable for applicati ons where speed is critical. T5, on the other hand, showed so lid performance in general content generation tasks, wh ile ERNIE stood out as the most balanced model, thanks to its ability to integrate structured data and effectivel y handle complex contexts. However, it is important to consider that NLP algor ithms may generate incorrect or inconsistent content, esp ecially when dealing with ambiguous contexts or if the info rmation they base their responses on contains errors or bia ses. This risk arises from how models interpret context or pr ioritize certain keywords over a full understanding of the c ontent. Therefore, users must validate the generated inform ation before using it, particularly in academic applicati ons where accuracy and reliability are essential. ACKNOWLEDGMENT The authors express their gratitude to the Direcció n de Investigación de la Universidad Peruana de Ciencias  
Aplicadas for the support provided for this researc h work. REFERENCES  
[1] M. Moreira, “Los materiales educativos: origen y fu turo,” 2007. [2] F. Gareis, F. Kendziur, & C. Machiavello, “Los desa fíos de la comunicación y la producción de materiales educativ os para Atamá, el ambiente virtual de aprendizaje de Entre Ríos,” 201 7. [3] C. Gamrat, L. Lenze, J. Bardzell, & E. Glantz, “Imp roved Student Engagement in Higher Education’s Next Normal,” 2021 . [4] P. M. Barbetta, “Tecnologías de escritura correctiv as y compensatorias para estudiantes de secundaria con problemas de apr endizaje,” 2023. [5] J. Franganillo, “La inteligencia artificial generat iva y su impacto en la creación de contenidos mediáticos,” 2023. [6] T. Farrely & N. Baker, “Generative Artificial Intel ligence: Implications and Considerations for Higher Education Practice,” 2023. [7] Y. Wenpeng, K. Katharina, & Y. Mo, “Comparative Stu dy of CNN and RNN for Natural Language Processing,” 2024. [8] A. Jain & A. Goel, “for natural language processing models\*,” 2020. [9] Y. Liu, M. Ott, N. Goyal, et al., “RoBERTa: A Robus tly Optimized BERT Pretraining Approach,” 2019. [10] J. Devlin et al., “BERT: Pre-training of Deep Bidir ectional Transformers for Language Understanding,” 2019. [11] A. Radford et al., “GPT-2: Better Language Models a nd Their Implications,” 2019. [12] M. Araujo, P. Gonçalves, & F. Benevenuto, “Towards understanding the impact of machine learning datasets on NLP task s,” 2020. [13] T. Brown et al., “Language models are few-shot lear ners,” 2020. [14] J. Devlin, M.-W. Chang, K. Lee, & K. Toutanova, “BE RT: Pre-training of deep bidirectional transformers for language und erstanding,” 2019. [15] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhut dinov, & Q. V. Le, “XLNet: Generalized autoregressive pretraining for language understanding,” 2019. [16] Y. Sun, S. Wang, Y. Li, S. Feng, & H. Tian, “ERNIE: Enhanced Representation through Knowledge Integration,” 2020 . [17] J. M. Swales & C. B. Feak, “Academic Writing for Gr aduate Students: Essential Tasks and Skills,” 2004. [18] M. McDonald, “Time Efficiency in Automated Text Gen eration Systems,” 2018  
Authorized licensed use limited to: Universidad Peruana de Ciencias Aplicadas (UPC). Downloaded on September 11,2025 at 04:08:42 UTC from IEEE Xplore. Restrictions apply.