IMPLEMENTING CONTINUAL LEARNING SCENARIOS: REVIEW OF A ML TECHNQIUE

AYSE ARSLAN

OXFORD ALUMNI F NORTEHR CALIFORNIA

DOI: https://doi.org/10.56293/IJASR.2022.5418

IJASR 2022
VOLUME 5
ISSUE 4 JULY – AUGUST **ISSN: 2581-7876**

Abstract: One of the biggest traps "AI" (artificial intelligence) projects fall into is demanding an entire solution fit into the ML (machine learning) paradigm. One of the continual learning techniques in ML to overcome this challenge is the so called 'Learning to Prompt for Continual Learning (L2P)' that can be applied to practical continual learning scenarios without known task identity or boundaries. L2P uses a single frozen backbone model and learns a prompt pool to conditionally instruct the model. After discussing main categories of recent continual learning algorithms, this paper provides an overview of LSP by discussing its layers.

Keywords: ML, AI, big data

INTRODUCTION

One of the biggest traps "AI" (artificial intelligence) projects fall into is demanding an entire solution fit into the ML (machine learning) paradigm. Such projects insist on learning everything from the data and discount any a priori knowledge of the problem. Consequently, unstable solutions are created on purpose by using the most complex solution for the most trivial aspects of a problem.

To overcome this challenge, this paper explores a continual learning technique - the so called 'Learning to Prompt for Continual Learning (L2P)' - that can be applied to practical continual learning scenarios without known task identity or boundaries. It provides a framework with a single frozen backbone model to display how it learns a prompt pool to conditionally instruct the model.

OVERVIEW OF EXISTING WORK

In 1950s, Alan Turing proposed a solution to the question of when a system designed by a human is 'intelligent.' Turing proposed a test that involves the capacity of a human listener to make the distinction of a conversation with a machine or another human-being; if this distinction is not detected, we can admit that we have an intelligent system, or artificial intelligence (AI).

In 1955, the term *artificial intelligence* was coined by the computer scientists McCarty, Minsky, Rochester, and Shannon as "computing systems that are able to engage in human-like processes such as learning, adapting, synthesizing, self-correction and use of data for complex processing tasks" (Popenici& Kerr, 2017, para. 3). Most AI applications involve the techniques of machine learning, deep learning, and natural language processing, among others.

- *Machine learning* (ML) is a subdiscipline of AI that consists of learning algorithms that use available data sources to summarize certain phenomena and further identify patterns. ML systems can be trained or learn to build a predictive model through supervised classification or unsupervised clustering(Ciolacu et al., 2018).
- *Deep learning* is a type of ML technology that uses artificial neural networks through layers of interconnected nodes to simulate the operation of the human brain(Cheng et al., 2018).
- *Natural language processing* is a field of AI related to understanding the human language through the analysis of sentences and the use of algorithms to extract the meaning of words. One well-known application is chatbots

that can understand common language requests and respond automatically, thus providing immediate assistance to users(Lu et al., 2020).

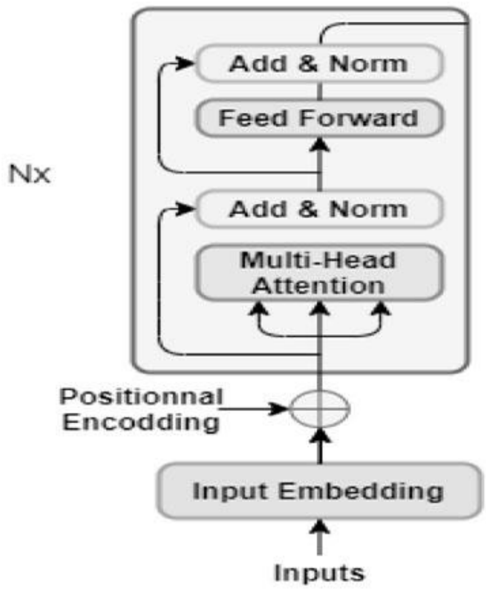
While ML can provide good solutions for many problems, it can't be perfect. Probabilistic outcomes can be considered as both a bug and a feature. It is a feature because probabilistic outcomes are more robust yet, it is also a bug as probabilistic outcomes can produce false positives, false negatives, and inconsistent value propositions for users.

Over the last several years, the rapidly growing size of deep learning models has quickly exceeded the memory capacity. While previous models like BERT can efficiently scale by leveraging data parallelism in which model weights are duplicated across accelerators while only partitioning and distributing the training data, recent large models like GPT-3 can only scale using model parallel training, where a single model is partitioned across different devices.

With only one line of code, any neural network can be transformed into a distributed version with an optimal parallelization strategy that can be executed on a user-provided device cluster. Existing ML parallelization strategies such as inter-operator parallelism and intra-operator parallelism assigns distinct operators to different devices which includes data parallelism, operator parallelism and expert parallelism so that collective communication is used to synchronize the results across devices.

The BERT model is based on two stages: pre-training and fine-tuning (Devlin et al., 2019). During pre-training, the model is trained on a large unlabeled corpus. The model is then fine-tuned, starting with the pre-trained parameters and refining all parameters with task-specific labeled data.

Figure 1. The BERT Encoder



BERT is a leading model for a variety of Natural Language Processing (NLP) tasks, demonstrating its efficiency and potential.

BERT is based on the attention mechanism (Vaswani et al., 2017) that was invented to allow a model to comprehend and remember the contextual relationships between features and text. BERT represents a single sentence or a pair of sentences as a sequence of tokens with the following characteristics:

- The first token in the sequence is [CLS].
- When there is a pair of sentences in the sequence, they are separated by the token [SEP].
- For a given token, its input representation is constructed by summing the corresponding token, position, and segment embeddings (see Figure 2).

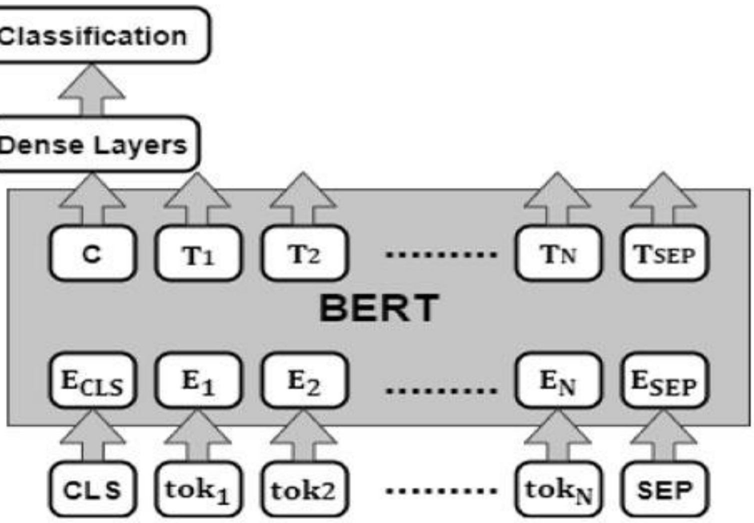


Figure 2. BERT Architecture

Despite theselimitations of hardware and software, AI providers could still benefit from a broader and more complex view of codes and customization by finding AI techniques and tools that are able to analyze complex codes and share an expanded view of both coding and functional patterns.

Advances in reinforcement learning (RL) have enabled agents to perform increasingly complex tasks in challenging environments.The notion of "value" in RL is intrinsically linked to *affordances*, in that the value of a state for skill reflects the probability of receiving a reward for successfully executing the skill. For an agent to perform complex tasks in realistic environments, it must be able to effectively reason over long horizons. The nature of the ideal state abstraction is closely tied to the action abstraction, as the most suitable abstraction of state should depend on the kinds of decisions that the higher-level policy needs to make, which in turn depends on the actions (skills) available to it.

For any skill, its value function captures two key properties: 1) the preconditions and affordances of the scene, i.e., where and when the skill can be used, and 2) the outcome, which indicates whether the skill executed successfully when it was used.

Moreover, Graph Neural Networks (GNNs) have extended the benefits of RL to the non-Euclidean domain, allowing for standardized and re-usable machine learning approaches to problems that involve relational (graph-structured) data [48]. GNNs are powerful machine learning (ML) models for graphs that leverage their inherent connections to incorporate context into predictions about items within the graph or the graph as a whole. A popular task for GNNs is node classification, in which a GNN is trained to infer node labels that represent some unknown property of each node, such as user interests in a social network.

Given a decision process with a finite set of k skills trained with sparse outcome rewards and their corresponding value functions, one can obtain an abstract representation that maps a state to a k-dimensional representation which captures *functional information* about the exhaustive set of interactions that the agent can have with the environment, and is thus a suitable state abstraction for downstream tasks.

When the relationship between input and output data in the underlying problem changes, ML will become obsolete over time. This is cause for diligent monitoring but also making a solution narrow. ML is also sensitive to the context or the background distribution of data. Such complex problems also arise due to changes in dataoccurrence, and could further be complicated by the context surrounding the data. To overcome the complexity, artificial neural networks (ANNs) which entail optimal features can be used to conduct excellent prediction and classification analysis.

Before transformers, recurrent neural networks (RNN) were the go-to solution for natural language processing. Numerous efforts have since continued to push the boundaries of recurrent language models and encoder-decoder architectures [38, 24, 15]. Recurrent models typically factor computation along the symbol positions of the input and output sequences. As RNN had to process data sequentially, they could not handle long sequences of text. Also, they only captured the relations between a word and the words that came before it. Transformers made it possible to process entire sequences in parallel, and to track the relations between words across very long text sequences in both forward and reverse directions.

When it comes to utilizing ML algorithms for form-based documents, it can be a challenge to convert unstructured text data into structured information as form documents often have more complex layouts that contain structured objects, such as tables, columns, and text blocks. Interwoven columns, tables, and text blocks make serialization difficult, substantially limiting the performance of a strict serialization approach.

A natural approach to handle form document understanding tasks is to first serialize the form documents (usually in a left-to-right, top-to-bottom fashion) and then apply state-of-the-art sequence models to them. To that end, Extended Transformer Construction (ETC; Ainslie et al., 2020) is a technique which scales transformers to long sequences with a sparse global-local attention mechanism. The aim is to mitigate the sub-optimal serialization of forms for document information extraction.

Most of the work done in ML has focused on supervised algorithms. Their main strength is that they produce models that can be incorporated in the decision-making process [4]. The ordinary supervised learning techniques use independent and identically distributed (IID) data, where all training examples are sampled from a fixed set of classes, and the model has access to these examples throughout the entire training phase.

One of these continual learning techniques in ML is the so called 'Learning to Prompt for Continual Learning (L2P)' thatcan be applied to practical continual learning scenarios without known task identity or boundaries. The next section will discuss the features in more detail.

Conceptual Framework

L2P leverages the representative features from pre-trained models; however, instead of tuning the parameters during the continual learning process, L2P keeps the pre-trained model untouched, and instead learns a set of prompts that dynamically instruct models to solve corresponding tasks.

There are three main categories of recent continual learning algorithms (Wang et. al, 2022):

- Regularization-based methods limit the plasticity of the model by limiting the learning rate on important parameters for previous tasks.
- Rehearsal-based methods construct a data buffer to save samples from older tasks to train with data from the current task.
- Architecture-based methods aim at having separate components for each task.

As it is displayed in Figure 3, L2P uses a single frozen backbone model and learns a prompt pool to conditionally instruct the model. "Model 0" indicates that the backbone model is fixed at the beginning (Wang et. al, 2022).

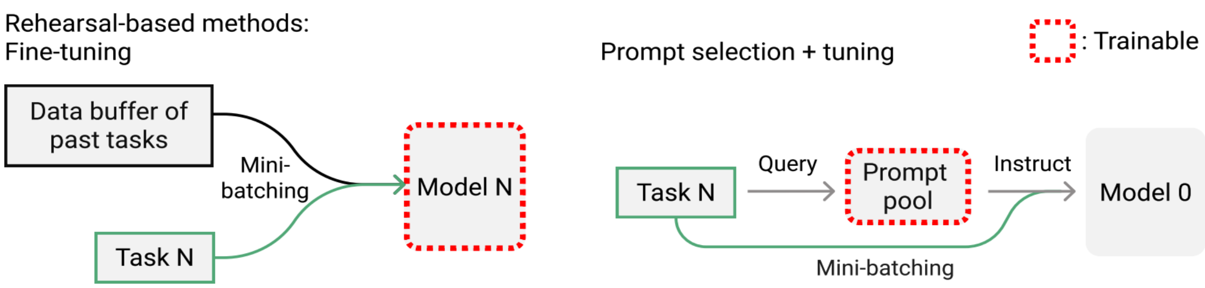


Figure 3. Overview of L2P

Given a pre-trained Transformer model, "prompt-based learning" modifies the original input using a fixed template.

Prompts that are frequently shared encode more generic knowledge while other prompts encode more task-specific knowledge.

This type of transformer architecture has evolved as the classic feed-forward neural network is not designed to keep track of sequential data and maps each input into an output. A ML model that processes text must not only compute every word but also take into consideration how words come in sequences and relate to each other. Here is an overview of problems that sequential neural networks purports to solve:

- A "vector to sequence" model takes a single input, such as an image, and produces a sequence of data, such as a description.
- A "sequence to vector" model takes a sequence as input, such as a product review or a social media post, and outputs a single value, such as a sentiment score.
- A "sequence to sequence" model takes a sequence as input, such as an English sentence, and outputs another sequence, such as the French translation of the sentence.

In a traditional attention layer, each token representation is linearly transformed into a Query vector, a Key vector, and a Value vector. At a high level, for each attention head at each layer, the aim is to examine each pair of token representations, to determine the ideal features the tokens should have if there is a meaningful relationship between them, and to penalize the attention score according to how different the actual features are from the ideal ones. This allows the model to learn constraints on attention using logical implication.

A token "looks" for other tokens from which it might want to absorb information (i.e., attend to) by finding the ones with Key vectors that create relatively high scores when matrix-multiplied by its Query vector and then softmax-normalized. The token then sums together the Value vectors of all other tokens in the sentence, weighted by their score, and passes this up the network, where it will normally be added to the token's original input vector.
Aligning the positions to steps in computation time, they generate a sequence of hidden states, as a function of the previous hidden state and the input for position t. The Transformer relies on an attention mechanism to draw global dependencies between input and output.

The input text must be processed and transformed into a unified format before being fed to the transformer. First, the text goes through a "tokenizer," which breaks it down into chunks of characters that can be processed separately. The tokens are then converted into "word embeddings." A word embedding is a vector that tries to capture the value of words in a multi-dimensional space. Word embeddings are created by embedding models, which are trained separately from the transformer. There are several pre-trained embedding models that are used for language tasks.

Once the sentence is transformed into a list of word embeddings, it is fed into the transformer's encoder module. It can receive an entire sentence's worth of embedding values and process them in parallel. This makes transformers more compute-efficient than their predecessors and also enables them to examine the context of the text in both forward and backward sequences.
The output of the attention layer is fed to a feed-forward neural network that transforms it into a vector representation and sends it to the next attention layer. The task of the decoder module is to translate the encoder's attention vector into the output data (e.g., the translated version of the input text). During the training phase, the decoder has access both to the attention vector produced by the encoder and the expected outcome (e.g., the translated string).The more training data and parameters the transformer has, the more capacity it gains to maintain coherence and consistency across long sequences of text.

Discussion

This study explored the use of a 'Learning to Prompt for Continual Learning (L2P)' that can be applied to practical continual learning scenarios without known task identity or boundaries. It expored how L2P uses a single frozen backbone model and learns a prompt pool to conditionally instruct the model.

An ongoing critique and inquiry in proposed solutions are critical to promote, and develop knowledge and wisdom. The same holds true for the field of AI. Despite a plethora of techniques, there is a need for research on the privacy implications of the current control on developments of AI with a focus on imagination, creativity, and innovation; the set of abilities and skills that can hardly be ever replicated by machines.

REFERENCES

1. Amisha, P. M., Pathania, M., &Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. Journal of Family Medicine and Primary Care, 8(7), 2328–2331.
2. Bedregal-Alpaca, N., Tupacyupanqui-Jaén, D., & Cornejo-Aparicio, V. (2020). Análisis del rendimientoacadémico de los estudiantes de Ingeniería de Sistemas, posibilidades de deserción y propuestas para suretención [Analysis of the academic performance of Systems Engineering students, dropout possibilities and proposals for their retention]. Ingeniare. RevistaChilena de Ingeniería, 28(4), 668–683. https://doi.org/10.4067/S0718-33052020000400668.
3. Bojorque, R., &Pesántez-Avilés, F. (2020) Academic quality management system audit using artificial intelligence techniques. In T. Ahram (Ed.), Advances in artificial intelligence, software and systems engineering (AHFE 2019) (Vol. 965, pp. 275–283). Springer. https://doi.org/10.1007/978-3-030-20454-9_28.
4. Bridgstock, R., & Jackson, D. (2019). Strategic institutional approaches to graduate employability: Navigating meanings, measurements and what really matters. Journal of Higher Education Policy and Management, 41(5), 468–484.
5. Brooks, C., & Thompson, C. (2017). Predictive modelling in teaching and learning. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), Handbook of learning analytics (pp. 61–68). Society for Learning Analytics Research.
6. Castro, D., Rodríguez-Gómez, D., &Gairín, J. (2017). Exclusion factors in Latin American higher education: A preliminary analyze from university governing board perspective. Education and Urban Society, 49(2), 229–247.
7. Cheng, F., Zhang, H., Fan, W., & Harris, B. (2018). Image recognition technology based on deep learning. Wireless Personal Communications, 102(2), 1917–1933.
8. Choque-Díaz, M., Armas-Aguirre, J., &Shiguihara-Juárez, P. (2018). Cognitive technology model to enhanced academic support services with chatbots. In Proceedings 2018 IEEE XXV international conference on electronics, electrical engineering and computing (INTERCON) (pp. 1–4). IEEE.
9. Ciolacu, M., Tehrani, A. F., Binder, L., &Svasta, P. M. (2018, October). Education 4.0-artificial intelligence assisted higher education: Early recognition system with machine learning to support students' success. In Proceedings 2018 IEEE 24th international symposium for design and technology in electronic packaging (SIITME) (pp. 23–30). IEEE.
10. Conley, C. S., Shapiro, J. B., Kirsch, A. C., &Durlak, J. A. (2017). A meta-analysis of indicated mental health prevention programs for at-risk higher education students. Journal of Counseling Psychology, 64(2), 121–140.
11. Contreras, L. E., Fuentes, H. J., & Rodríguez, J. I. (2020). Predicción del rendimientoacadémicocomoindicador de éxito/fracaso de los estudiantes de ingeniería, medianteaprendizajeautomático [Predicting academic performance as an indicator of success/failure of engineering students, using machine learning]. FormaciónUniversitaria, 13(5), 233–246. https://doi.org/10.4067/S0718-50062020000500233.
12. Cordero, J., Toledo, A., Guamán, F., & Barba-Guamán, L. (2020). Use of chatbots for user service in higher education institutions. In Proceedings 2020 15th Iberian conference on information systems and technologies (CISTI) (pp. 1–6). IEEE. https://doi.org/10.23919/CISTI49556.2020.9141108.
13. da Fonseca Silveira, R., Holanda, M., de Carvalho Victorino, M., &Ladeira, M. (2019). Educational data mining: Analysis of drop out of engineering majors at the UnB—Brazil. In Proceedings 2019 18th IEEE international conference on machine learning and applications (ICMLA) (pp. 259–262). IEEE. https://doi.org/10.1109/ICMLA.2019.00048.
14. Dagli, G., Altinay, F., Altinay, Z., &Altinay, M. (2020). Evaluation of higher education services: Social media learning. The International Journal of Information and Learning Technology, 38(1), 147–159.
15. Delahoz-Dominguez, E. J., Guillen-Ibarra, S., &Fontalvo-Herrera, T. (2020). Análisis de la acreditación de calidadenprogramas de ingeniería industrial y los resultadosen las pruebasnacionalesestandarizadas, en Colombia [Analysis of quality accreditation in industrial engineering programs and results in national

standardized tests, in Colombia]. Formaciónuniversitaria, 13(1), 127-134. https://doi.org/10.4067/S0718-50062020000100127.
16. de Peña, L. P., & Pérez, A. M. C. (2013). Review of some studies on university student dropout in Colombia and Latin America. Acta Universitaria, 23(4), 37–46.
17. Dehon, P., Silva, A., Inocêncio, A. C., Castro, C., Costa, H., & Júnior, P. P. (2018, October). CVchatbot: Um chatbot para o aplicativo Facebook Lessengerintegradoao AVA Moodle [CVChatbot: A chatbot for the Facebook Messenger application integrated with LMS Moodle]. In Proceedings Brazilian symposium on computers in education (SimpósioBrasileiro de InformáticanaEducação-SBIE) (Vol. 29, No. 1, pp. 1623–1632). https://doi.org/10.5753/cbie.sbie.2018.1623
18. Dzhangarov, A. I., Suleymanova, M. A., &Zolkin, A. L. (2020, May). Face recognition methods. In IOP conference series: Materials science and engineering (Vol. 862, No. 042046). IOP Publishing. https://doi.org/10.1088/1757-899X/862/4/042046
19. Erkan, A. (2019). Impact of using technology on teacher-student communication/interaction: Improve students learning. World Journal of Education, 9(4), 30–40.
20. Espinosa Rodríguez, R., Pérez de Celis Herrero, C., del CarmenLara Muñoz, M., Somodevilla García, M. J., & Pineda Torres, I. H. (2018). Chatbots en redes sociales para el apoyooportuno de estudiantesuniversitarios con síntomas de trastorno por déficit de la atención con hiperactividad [Social media chatbots for timely support of college students with symptoms of attention deficit hyperactivity disorder]. RevistaIberoamericana de TecnologíaenEducación y EducaciónenTecnología, 22, 52–62. https://doi.org/10.24215/18509959.22.e06.
21. Fiallos, A., Jimenes, K., Vaca, C., & Ochoa, X. (2017). Scientific communities detection and analysis in the bibliographic database: SCOPUS. In Proceedings 2017 fourth international conference on eDemocracy& eGovernment (ICEDEG) (pp. 118–124). IEEE. https://doi.org/10.1109/ICEDEG.2017.7962521.
22. Gade, K., Geyik, S., Kenthapadi, K., Mithal, V., &Taly, A. (2020, April). Explainable AI in industry: Practical challenges and lessons learned. In A. E. F. Seghrouchni, G. Sukthankar, & T.-Y. Liu (Eds.), Companion proceedings of the web conference 2020 (pp. 303–304). ACM.
23. *García-González, D. J., &Skrita, A. (2019). Predicting academic performance based on students' family environment: Evidence for Colombia using classification trees. Psychology Society & Education, 11(3), 299–311. https://doi.org/10.21071/psye.v11i3.13947.
24. García-Vélez, A. R., López-Nores, M., González-Fernández, G., Robles-Bykbaev, V. E., Wallace, M., Pazos-Arias, J. J., & Gil-Solla, A. (2019). On data protection regulations, big data and sledgehammers in higher education. Applied Sciences-Basel, 9(15), 3084. https://doi.org/10.3390/app9153084
25. Gómez, L. F., & Valdés, M. G. (2019). The evaluation of teacher performance in higher education. Journal of Educational Psychology-Propósitos y Representaciones, 7(2), 499–515. https://doi.org/10.20511/pyr2019.v7n2.255.
26. Gómez Cravioto, D. A., Díaz Ramos, R. E., ZenklGalaz, M. A., Hernández Gress, N., & Ceballos Cancino, H. G. (2020). Analysing factors that influence alumni graduate studies attainment with decision trees. In Proceedings 2020 international conference on computer science and software engineering (CSASE) (pp. 44–49). IEEE. https://doi.org/10.1109/CSASE48920.2020.9142069.
27. Gutiérrez, G., Canul-Reich, J., Ochoa Zezzatti, A., Margain, L., & Ponce, J. (2018). Mining: Students comments about teacher performance assessment using machine learning algorithms. International Journal of Combinatorial Optimization Problems and Informatics, 9(3), 26–40. https://ijcopi.org/ojs/article/view/99
28. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. Knowledge-Based Systems, 161, 134–146.
29. Khairil, L. F., &Mokshein, S. E. (2018). 21st century assessment: Online assessment. International Journal of Academic Research in Business and Social Sciences, 8(1), 659–672.
30. Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—A systematic literature review. Information and Software Technology, 51(1), 7–15.
31. *Klos, M. C., Escoredo, M., Joerin, A., Lemos, V. N., Rauws, M., & Bunge, E. L. (2021). Artificial intelligence chatbot for anxiety and depression in university students: A pilot randomized controlled trial. JMIR Formative Research, 5(8), e20678. https://doi.org/10.2196/20678
32. Kumar, V. (2017). The role of university centers in promoting research. Journal of the Academy of Marketing Science, 45(4), 453–458.

33. Longo L. (2020). Empowering qualitative research methods in education with artificial intelligence. In A. Costa, L. Reis, & A. Moreira (Eds.), Computer supported qualitative research (Vol. 1068, pp. 1–21). Springer.
34. Lu, X. (2018). Natural language processing and intelligent computer-assisted language learning (ICALL). In J. I. Liontas (Ed.), The TESOL encyclopedia of English language teaching (pp. 1–6). Wiley Blackwell.
35. Lu, X., Li, S., & Fujimoto, M. (2020). Automatic speech recognition. In Y. Kidawara, E. Sumita, & H. Kawai (Eds.), Speech-to-speech translation (pp. 21–38). Springer.
36. *Mendoza Jurado, H. F. (2020). Modelos de redes neuronalesartificiales, comosustentoevaluativo al crecimientopedagógico virtual enEducación Superior [Artificial neural network models, as evaluative support to virtual pedagogical growth in Higher Education]. Educación Superior, 7(2), 25–36. https://www.scielo.org.bo/pdf/escepies/v7n2/v7n2_a06.pdf.
37. *Menezes, A. G., da C. Sá, J. M. D., Llapa, E., &Estombelo-Montesco, C. A. (2020). Automatic attendance management system based on deep one-shot learning. In Proceedings 2020 international conference on systems, signals and image processing (IWSSIP) (pp. 137–142). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145230.
38. Miranda, M. A., & Guzmán, J. (2017). Análisis de la deserción de estudiantesuniversitariosusandotécnicas de minería de datos [Analysis of university student dropout using data mining techniques]. FormaciónUniversitaria, 10(3), 61–68. https://doi.org/10.4067/S0718-50062017000300007.
39. *Nieto, Y., García-Díaz, V., Montenegro, C., González, C. C., & González Crespo, R. (2019). Usage of machine learning for strategic decision making at higher educational institutions. IEEE Access, 7, 75007–75017. https://doi.org/10.1109/ACCESS.2019.2919343.
40. Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Hammout, N., Nakamura, E. L., Escamilla, J., & Hosseini, S. (2020). Impact of students' evaluation of teaching: A text analysis of the teachers qualities by gender. International Journal of Educational Technology in Higher Education, 17(1), 1–27. https://doi.org/10.1186/s41239-020-00224-z.
41. Owoc M. L., Sawicka A., &Weichbroth P. (2021). Artificial intelligence technologies in education: Benefits, challenges and strategies of implementation. In M. L Owoc, & M. Pondel (Eds.), Artificial intelligence for knowledge management (Vol. 599, pp. 37–58). Springer.
42. Padayachee, P., Wagner-Welsh, S., & Johannes, H. (2018). Online assessment in Moodle: A framework for supporting our students. South African Journal of Higher Education, 32(5), 211–235.
43. Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. Entropy, 23(4), 485. https://doi.org/10.3390/e23040485
44. Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning. https://doi.org/10.1186/s41039-017-0062-8
45. Salas-Pilco, S. Z., & Yang, Y. (2020). Learning analytics initiatives in Latin America: Implications for educational researchers, practitioners and decision makers. British Journal of Educational Technology, 51(4), 875–891. https://doi.org/10.1111/bjet.12952
46. Santos, G. A. S., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., &Brandao, D. N. (2020). EvolveDTree: Analyzing student dropout in universities. In Proceedings 2020 international conference on systems, signals and image processing (IWSSIP) (pp. 173–178). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145203.
47. Sayama, H. F., Araujo, A. V., & Fernandes, E. R. (2019). FaQuAD: Reading comprehension dataset in the domain of Brazilian higher education. In Proceedings 2019 8th Brazilian conference on intelligent systems (BRACIS) (pp. 443–448). IEEE. https://doi.org/10.1109/BRACIS.2019.00084.
48. Schwartzman, S. (Ed.) (2020). Higher education in Latin America and the challenges of the 21st century. Springer. Siemens, G. (2019). Learning analytics and open, flexible, and distance learning. Distance Education, 40(3), 414–418.
49. Sun, Z., &Stranieri, A. (2021). The nature of intelligent analytics. In Z. Sun (Ed.), Intelligent analytics with advanced multi-industry applications (pp. 1–21). IGI Global.
50. The World Bank. (2021b). Latin America and the Caribbean: Tertiary education. Context. http://pubdocs.worldbank.org/en/738931611934489480/LAC-TE-and-Covid-updated.pdf
51. The World Bank. (2021a). World Bank open data. Retrieved July 5, 2021, from https://data.worldbank.org/

52. Torre, D., & Zapata, G. (2012). Impacto de procesos de aseguramiento de la calidadsobre las instituciones de educación superior: un estudioensietepaíses [Impact of quality assurance processes on higher education institutions: A study in seven countries]. In M. Lemaitre, & M. Zenteno (Eds.), Aseguramiento de la Calidad enIberoamérica. Informe 2012 (pp. 115–157). RIL Editores.
53. *Torres Soto, M. D., Torres Soto, A., Barajas Aranda, D. A., Campos Muñoz, N., Ponce de León Sentí, E. E., & Velazquez Amador, C. E. (2019). Suicidal tendency neural identifier in university students from Aguascalientes, Mexico. In M. A. Carreno León, J. A. Sandoval Bringas, M. Chacón Rivas, F. J. Álvarez Rodríguez, & A. Silva Sprock (Eds.), Proceedings 2019 XIV Latin American conference on learning technologies (LACLO) (pp. 387–392). IEEE. https://doi.org/10.1109/LACLO49268.2019.00071.
54. Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C. C. (2017). Some guidance on conducting and reporting qualitative studies. Computers & Education, 106, A1–A9.
55. *Ulloa Cazarez, R. L., & López Martín, C. (2018). Neural networks for predicting student performance in online education. IEEE Latin America Transactions, 16(7), 2053–2060. https://doi.org/10.1109/TLA.2018.8447376.
56. *Villaseñor, A. E., Arencibia-Jorge, R., & Carrillo-Calvet, H. (2017). Multiparametric characterization of scientometric performance profiles assisted by neural networks: A study of Mexican higher education institutions. Scientometrics, 110(1), 77–104. https://doi.org/10.1007/s11192-016-2166-0.
57. Visbal-Cadavid, D., Mendoza, A. M., &Hoyos, I. Q. (2019). Prediction of efficiency in colombian higher education institutions with data envelopment analysis and neural networks. PesquisaOperacional, 39(2), 261–275. https://doi.org/10.1590/0101-7438.2019.039.02.0261.
58. Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of Planning Education and Research, 39(1), 93–112.
59. Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. Computers and Education: Artificial intelligence, 2, 100025.

