Epistemic neural network based evaluation of online teaching status during epidemic period

Ni Yao

Received: 27 June 2022 / Revised: 19 September 2022 / Accepted: 15 October 2022 / Published online: 8 November 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
During the epidemic, online teaching became the mainstream. Online teaching evaluation aims to systematically test teachers' teaching process according to certain teaching objectives and standards, and evaluate its value, advantages and disadvantages, so as to improve the quality of teaching. It is not only an important part of the teaching process, but also the basis of all effective and successful teaching. In this paper, we propose an online teaching evaluation method based on Epistemic Neural Network (ENN), which is an evolutionary intelligence method. In terms of uncertainty modeling, ENN's design innovation provides the improvement effect of geometric progression in terms of statistical quality and calculation cost. Therefore, it is very suitable for teaching evaluation, which is an evaluation process guided by a variety of uncertain factors. Specifically, this paper considers the content and grade standards of online teaching evaluation from five aspects. (1) Teachers' syllabus, teaching progress, teaching plan, courseware and other teaching documents and teaching materials; (2) Abide by teaching discipline, the implementation of teaching plan and the completion of teaching tasks; (3) Teaching attitude, teaching investment, teaching and educating people, and the comprehensive quality of teachers; (4) Whether the concepts taught in the course are accurate, the expression is clear, whether the key points are prominent and whether the difficulties are clearly explained; (5) The depth, breadth and frontier of teaching content, and the amount of classroom information. According to the above five evaluation indexes which involves the big data analysis, we train ENN to get an evaluation score that can evaluate the teacher's online teaching process. In addition, we also test the average evaluation time to verify the effectiveness.

Keywords Online teaching · ENN · Evolutionary intelligence · Data analysis

1 Introduction
Teaching status evaluation is an arduous and difficult but significant teaching task. It is also an effective means to ensure the quality of higher education in many countries in the world. Developed countries have carried out higher education evaluation for decades or even hundreds of years, and pay more and more attention to evaluation. They regard evaluation as an important measure to strengthen macro-control and improve the quality of education. In addition, practice has also proved that the evaluation work has played a very important role in promoting colleges and universities to change educational ideas, establish modern educational concepts, strengthen teaching work, implement the central position of teaching work, improve school running conditions, standardize teaching management and improve education quality, which has been fully affirmed and highly praised by the higher education front and all aspects of society. It is also affected by the transition from online assessment to offline teaching. Compared with offline teaching evaluation, online teaching evaluation can solve the defects of the existing evaluation work through technical means. For example, the teaching conditions and management of some colleges and universities are not good enough, which directly affect the improvement of education quality; In order to cope with the assessment, some schools organize teachers and students to make up test papers and papers, and even fabricate false documents and data. They also attribute this kind of fraud to the assessment. It must be noted that it is not the assessment that leads to fraud, but through the assessment, the problems existing in school management and the problems existing in the thought and integrity of school leaders are exposed. All the above problems are caused by the strong subjective
orientation of the existing teaching evaluation. Therefore, there is an urgent need for a more objective teaching evaluation through technical means.

Neural network is the most popular machine learning method. As an objective technology, it can be used as a means of online teaching evaluation. The online teaching evaluation process usually requires effective decision-making and exploration. The evaluator needs to know what he knows and what he doesn't know. This capability depends on the quality of the joint prediction of tags assigned to multiple inputs. Traditional neural networks lack this ability, and because most studies focus on edge prediction, this defect is largely ignored. By evaluating the quality of joint prediction, it can be determined whether the neural network can effectively distinguish between cognitive uncertainty (due to lack of knowledge) and arbitrary uncertainty (due to contingency). Epistemic Neural Network (ENN) [1] was proposed by DeepMind in 2021 and can be used to solve this problem. ENN can be used as an interface for uncertainty modeling in deep learning. All existing uncertainty modeling methods can be expressed as ENN, and any ENN can be constructed by Bayesian neural network. Moreover, by introducing epinet, ENN can supplement any existing neural network architecture, including pre-training model, and train through moderate incremental calculation to represent uncertainty. Using epinet, the traditional neural network output forms a very large set, which is composed of hundreds or more particles with less computation.

Based on the above basic ENN architecture, we select five aspects of online teaching evaluation content and grade standards, namely, (1) teachers' syllabus, teaching progress, teaching plan, courseware and other teaching documents and teaching materials; (2) Abide by teaching discipline, the implementation of teaching plan and the completion of teaching tasks; (3) Teaching attitude, teaching investment, teaching and educating people, and the comprehensive quality of teachers; (4) Whether the concepts taught in the course are accurate, the expression is clear, whether the key points are prominent and whether the difficulties are clearly explained; (5) The depth, breadth and frontier of teaching content, and the amount of classroom information. The whole process considers five evaluation indexes, which requires the big data analysis. Therefore, this paper needs to train ENN and finally get an evaluation score that can evaluate the teacher's online teaching process.

The rest paper is organized as follows. Section 2 reviews the related work; Sect. 3 presents the technical background; Sect. 4 proposed ENN-based method; Sect. 5 reports the experimental results; Sect. 6 concludes this paper.

2 Related work

2.1 Traditional online teaching evaluation methods

The traditional methods for online teaching evaluation can be classified into two folds. (1) Identify teachers' course design, evaluation and assessment, and facilitation practices from the methods of award-winning online faculty. Martin et al. [2] developed a conceptual framework to identify teachers' course design. They interviewed 8 award-winning online faculty members who received online teaching awards from online learning consortium and association educational communications. Ibrahim et al. [3] solved the online teaching evaluation problem of universities in Jordan by evaluating the online teaching of architecture design. Their study organized a unique group meeting of 10 expert persons, then a questionnaire survey including faculty members and a randomly students were visited. The survey results presented that online teaching during pandemic was regularly in synchronous meetings format and faculties and students were met with this type of theoretical courses but did not satisfy the basic design courses. Jones et al. [4] solved social distancing problem of face-to-face teaching in universities under the COVID-19 pandemic scene. They studied student feedback to a course in detail which indicated whether an applied online course would be well received. (2) Consider evaluation of online teaching quality as a problem of linguistic multi-attribute group decision-making [5]. Lin et al. [5] considered that the evaluation semantic information was distributed either symmetrically or asymmetrically in a linguistic term set and they extended linguistic framework to evaluate online teaching quality by using both the risk preferences of assessment experts and unknown weight information of attributes and sub-attributes. This method maximized teachers’ group comprehensive rating values, so as to evaluate the overall teaching quality.

2.2 Machine-learning-based online teaching evaluation methods

Recently, many scholars utilizes support vector machine and deep neural network to evaluation the quality of online teaching [6–8]. Li et al. [6] evaluated online teaching qualities of basic education based on artificial intelligence. They proposed an entropy weight method and introduced a gray clustering analysis to evaluate the online teaching quality in basic education. The authors also gave several strategies to improve the quality of online teaching in basic education on the basis of their proposed model. They claimed that the research results provided a good reference...
for AI-based online teaching in basic education. Hou et al. [7] considered that the recent online education evaluation models were inadequate in solving small-data scale sets of evaluation, thus further research machine learning methods for online teaching. They introduced adaptive learning rate and momentum terms to help the gradient descent method of BP neural network to accelerate the convergence speed of the model. In addition, their proposed deep neural network model was claimed to be capable in dealing with complex, high-dimensional, and large-scale data sets. Their results presented that the method was effective and advantageous in evaluating teaching quality in universities where there were usually large-scale data sets. Jiang et al. [8] proposed an online teaching quality evaluation machine learning model based on analytic hierarchy process and particle swarm optimization BP neural network. In their method, by using the analytic hierarchy process, an online teaching quality evaluation system was constructed to set the weight of the machine learning model; according to safety regulations, the risk value of each index was constructed with combining the actual experience. The authors optimized a BP neural network model that was subsequently used to be a regression model. The optimized algorithm was the particle swarm. The parameters of the BP model could be constantly adjusted, where the appropriate hyper-parameters were selected including the particle swarm algorithm. Their experimental results and comparison results presented that the accuracy of the optimized model was higher than that of the baseline BP model and could effectively overcome the drawbacks of BP neural network.

Although the above-mentioned methods addressed the online teaching problem greatly, the education problem always needs to be optimized. Different from them, the proposed evaluation method has three outstanding points. At first, this paper considered five comprehensive evaluation indexes; Then, this paper used ENN which was an evolutionary intelligence method to process evaluation indexes. Finally, this paper realized the unification of data forms during feature extraction.

3 Preliminaries

In this section, we introduce some basic knowledge about machine learning, neural network, and ENN.

3.1 Machine learning

Machine learning [9–12] is an interdisciplinary subject involving probability theory, statistics, approximation theory, convex analysis, algorithm complexity theory and so on. It specially studies how computers simulate or realize human learning behavior, so as to acquire new knowledge or skills, reorganize the existing knowledge structure and continuously improve its own performance. It is the core of artificial intelligence and the fundamental way to make computers intelligent. Machine learning has the following three definitions [13]: (1) Machine learning is a science of artificial intelligence. The main research object in this field is artificial intelligence, especially how to improve the performance of specific algorithms in empirical learning. (2) Machine learning is the study of computer algorithms that can be automatically improved by experience. (3) Machine learning is the use of data or past experience to optimize the performance criteria of computer programs.

3.2 Neural network

Neural network (NN) or fully connected model is an algorithm mathematical model that simulates the behavior characteristics of biological neural network and performs distributed parallel information processing. Depending on the complexity of the system, this network can process information by adjusting the connection between a large number of internal nodes [14]. Biological neural network mainly refers to the neural network of human brain, which is the technical prototype of artificial neural network. Biological neural network mainly studies the structure, function and working mechanism of human brain neural network, aiming to explore the law of human brain thinking and intelligent activities. Artificial neural network is the technical reproduction of biological neural network in a simplified sense. As a discipline, its main task is to build a practical artificial neural network model according to the principle of biological neural network and the needs of practical application, design the corresponding learning algorithm, simulate some intelligent activities of the human brain, and then realize it technically to solve practical problems. All in all, biological neural network mainly studies the mechanism of intelligence, and artificial neural network mainly studies the realization of intelligent mechanism, and the two complement each other [15, 16].

As shown in Fig. 1 [17], NN consists of node layer, including an input layer, one or more hidden layers and an output layer. Each node, also known as an artificial neuron, is connected to another node with associated weights and thresholds. If the output of any single node is higher than the specified threshold, the node will be activated and send data to the next layer of the network. Otherwise, the data will not be transferred to the next layer of the network. Therefore, in each node, there is an activation function. The typical activation functions are sigmoid function $y = \frac{1}{1+e^{-x}}$, tanh function $\tanh(x) = \frac{\cosh(x)}{\sinh(x)}$. By observing the above figure, we can find that its rules include: Neurons are arranged in layers.
The leftmost layer is called the input layer, which is responsible for receiving input data; The rightmost layer is called the output layer, from which we can obtain the output data of the neural network. The layers between the input and output layers are called hidden layers because they are invisible to the outside. There is no connection between neurons in the same layer. Each neuron in layer N is connected to all neurons in layer $n-1$ (this is the meaning of full connected). The output of neurons in layer $n-1$ is the input of neurons in layer $n$. Each connection has a weight. These rules define the structure of a fully connected neural network. In fact, there are many other neural networks, such as convolutional neural network (CNN) [18] and recurrent neural network (RNN) [19], which have different connection rules.

### 3.3 Epistemic neural network

ENN is proposed by Osband et al. [1]. Its network model including input, index, network architecture, and output are presented in Fig. 2. Given an input example $x$ and initial parameters $w$, a CNN produces an output $f_w(x, z)$. Specially, the output $f_w(x, z)$ of the ENN depends additionally on an epistemic index $z$. It assigns a regular reference distribution $P_Z$ for the index $z$. Typical choices of the distribution contain a uniform distribution over finite fields or a standard Gaussian over vector spaces. The index $z$ is used to indicate epistemic uncertainty. In particular, the variation of network output with $z$ indicates that future data may solve uncertainty. The introduction of the epistemic index supports us to express the uncertainty required to generate useful union predictions. In typical classification tasks, the traditional neural network generates a marginal prediction result and distributes a probability $\text{softmax}(f_w(x, z))$ to each data label category $y$. A union prediction results across the inputs $x_1, \ldots, x_c$ would distribute a probability to each category that combine $y_1, \ldots, y_c$. While CNNs are not provided for union predictions, it can be generated by jointing multiple marginal prediction results.

Although the ENN framework and notation are new, all of the existing uncertainty estimation methods can be expressed as an ENN. For example, the Bayesian neural network (BNN) architecture as shown in Fig. 3, can be easily written as an ENN. Where a CNN produces output $f_w(x, z)$, a BNN then maintains a posterior distribution over weights $w$ [20]. For any posterior distribution, we can define reference distribution $P_Z$ and parameterized class of functions $g_\theta$ such that a prediction $f_{g_\theta}(Z)(x)$, with $Z$ sampled from $P_Z$, reflects predictive uncertainty equivalent to that of the original BNN. Concretely, a ‘deep ensemble’ [21] can be expressed as an ENN by taking $Z$ to index particles of the ensemble and the reference distribution $P_Z$ to be uniform across these indices. For Monte Carlo dropout [22], $Z$ can be took to identify the dropout mask.

### 4 Method

In order to evaluate the online teaching, we divide the evaluation system into data acquisition module, data analysis module and teaching evaluation module. The data acquisition module firstly collects reference data related to teaching evaluation, including five aspects, i.e., (1) the teachers’ syllabus, teaching progress, teaching plan, courseware and other teaching documents and teaching materials; (2) teacher’s teaching plan and completion; (3) the teachers’ teaching attitude, teaching investment, teaching and educating people, and the comprehensive quality of teachers; (4) whether the concepts taught in the course are accurate, and the expression is clear, whether the key points are prominent and whether the difficulties are clearly explained; (5) as for the teaching depth, breadth and frontier of teaching content,
and the amount of classroom information. At the beginning of teaching evaluation, teachers only need to upload corresponding teaching documents and forms directly. After the implementation of teaching, students need to fill in a questionnaire, and the system will collect various test scores as the main reference data sources of teaching evaluation.

The second module is data analysis module. This module is set to transform the different forms data collected from the data acquisition module into the uniform perception forms. Therefore, we should extract these data features for teaching evaluation. Because these data have different modes, we use various data processing methods to process these modal data into uniform feature vector form. These technologies include natural language processing, speech recognition, text classification, emotional degree regression, etc. Following the multi-data perception module, we extract the embedded features of different evaluation elements respectively. Then, these features are feed to ENN for teaching evaluation.

The last module is the teaching evaluation module. This module analysis all the input teaching related data and output the final teaching score. This module adopts the structure of ENN and carries out epistemic training through labeled data. When the network converges, we save the network model parameters for subsequent teaching evaluation inference. The following sections will introduce the details of the above three modules.

4.1 Data acquisition

Our teaching evaluation system runs through the whole teaching activity. Teachers’ teaching outlines, teaching plans, teaching notes, online classes, classroom assignments, classroom assessments, teaching tests, after-class questionnaires and so on are all carried out on our online teaching system. Therefore, our teaching system ensures the integrity and real-time of the teaching evaluation data. By this way, our online teaching platform also acts as a data acquisition module, directly obtaining the teaching data of teachers with a specific ID through the background. Through database management, each teacher’s teaching data dictionary is established for subsequent teaching evaluation and improvement.

The whole teaching process covers all kinds of multimedia data, including text, video, voice and natural language. These multi-modal data are associated by a unified ID. The ID consists of 16 digits, with the first 4 indicating the teacher and the last 12 index the teaching sessions. In this way, we can associate multi-modal data, and at the same time, we can also associate current data with historical data, which helps us to make more objective and accurate teaching evaluation and evaluate the improvement of teaching tasks according to the historical data of teachers.

4.2 Data analysis

In order to make use of these multi-modal data for online teaching evaluation, we need to extract features from various forms of data, and realize the unification of data forms during feature extraction, that is, different data are unified into the form of data features. To do this, we use advanced artificial intelligence technology for multi-modal data analysis to extract high-level features, which is a high generalization of the literal meaning. For different data, we use different neural networks for feature extraction and fusion.

Specifically, we use the popular CLIP model [23] to process the video frames and the text data in the video and use it to implement the feature extraction and fusion. The videos include teachers’ teaching videos and students’ online learning videos. For some research data and examination results, there are a lot of scalar data, and vector data. In order to deeply explore the relationship between these data and teaching evaluation scores, we use the embedding method to transform the original numerical data into feature vectors of certain dimensions. For example, as for a scalar data, we use a $1 \times m$ Multilayer perceptron (MLP) to transform the data into a feature vector, where the input data has the size of $1 \times 1$, and the output of the MLP is a vector form with the size of $1 \times m$. The structure of the network is shown in Fig. 4a. For a vector data such as the teaching evaluation results with the dimension of $n$, we also use MLP to embed the

*Fig. 4* The scalar data and vector data are embedded into the uniform feature dimension
n-dimensional vector into a m-dimensional vector as shown in Fig. 4b. When the two feature dimensions are the same, the two features can be directly added and subtracted or cascaded. Through this method, we can transform all kinds of data related to online teaching evaluation into the same data form, which is beneficial for ENN to analyze these data and get teaching evaluation scores.

### 4.3 Teaching evaluation

After unifying all kinds of data into the same form of features, we use these features to regress the online teaching evaluation score. In this module, we propose a new teaching evaluation network based on the ENN model proposed by Osband [1]. Be different from the ENN model, in our model, we set up a branch of the reference feature network for historical information. Through the reference to the historical confidence, the evaluation can be more comprehensive and objective to the current teaching level of horizontal and vertical. In this paper, we use historical features and historical scores as two additional reference information for the current teaching evaluation. In order to avoid much influence of history on the current forecast, the weight of historical data is generally set to 0.2.

As shown in Fig. 5, which is our ENN based on the historical information reference. The original index \( z \) is formulated as the ID that indexes the teachers and the teaching sessions. The input multi-modal features are generated by the concatenated features of all the features extracted in the data analysis module. In our network structure, the indexed historical features are formulated as the reference in the ENN. Then the output is the online teaching evaluation score. In order to evaluate the current teaching improvement, we also input the ID indexed historical teaching evaluation score as another reference.

### 5 Implementation and experiment

Our main algorithms are implemented with the PyTorch deep learning framework. In order to train our network, we first collect and make a labeled dataset for training supervision with the traditional expert scoring system. These data are provided by the traditional school teaching evaluation system. We classify these traditional artificial evaluation data, establish the data structure of the above system, and carefully check and evaluate the objectivity, representativeness and diversity of teaching scores. After preparing the data, we used a server with 4 3090 GPUs to train the network for 2 days. Then we get a convergent model which can be used to evaluate online teaching directly by inference model. In order to make an objective evaluation of the performance of the evaluation system, we reserve some training data for network testing while making training data. Therefore, we can use these validation datasets to evaluate the model performance and feasibility of the proposed method.

The final experimental results show that the error between the evaluation score obtained by inference with this model and that obtained by previous evaluation system is within 5%. By carefully analyzing the scores predicted by the model, we find that the results obtained by this method are more objective and can effectively avoid the deviation of teaching evaluation scores caused by human factors. For example, the result of Table 1 shown that the subjective scores of the three semesters is very close due to subjective

| Test            | ENN | Subjective score | Comprehensive score | Error (%) |
|-----------------|-----|------------------|---------------------|-----------|
| The first semester | 75.3 | 77.5             | 76.2                | 1.2       |
| The second semester | 90.5 | 79.5             | 89.7                | 0.9       |
| The third semester  | 89.3 | 81.0             | 92.8                | 3.7       |

Table 1 The comparison between the artificial method and the ENN method
first impressions. In order to avoid subjective factors and get an accurate comprehensive score, we asked five different experts to score the average comprehensive score. The results show that the prediction result of our epistemic neural network is very close to the comprehensive score of five experts. This is because we kept the manual initial score while making objective score. Moreover, each teacher’s evaluation input is independent and fair, so it is more objective.

Furthermore, this paper tests the evaluation time to verify the effectiveness, where references [6–8] are considered as the baselines. In particular, the number of simulations is set as 10. The experimental results regarding the average evaluation time are shown in Fig. 6. It is observed that, the proposed ENN-based method has the smallest average evaluation time, only for 38.267 s, followed by Ref8, Ref7 and Ref6, this is because this paper has three outstanding points mentioned at the end of Sect. 2. In addition, Ref6 has the largest average evaluation time, about 61.921 s, because it did not consider the training process, which could not support the accurate analysis and thus spent much time.

6 Conclusion

In this paper, we propose a new online teaching status evaluation system based on ENN. The system provides more objective, accurate and university evaluations of online teaching models widely used during the pandemic period. The system uses the most advanced multi-modal data feature extraction methods and ENN for online teaching evaluation, which fully considers the evolutionary intelligence data analysis method. The online teaching mode caused by the epidemic made students and teachers not adapt to the phenomenon at the initial stage. In order to solve this problem, the online teaching evaluation method proposed by this paper effectively improves the quality of online teaching and speeds up the adaptation of students and teachers to this new education mode. This is of great significance to effectively improve students’ learning efficiency and teachers' teaching level during the epidemic period. However, the current methods still rely heavily on expert annotated data, which limits the application and promotion of the algorithm. Therefore, how to conduct self-supervised teaching evaluation is a further work to be carried out.

Declarations

Conflict of interest The author declares that there is no conflicts of interest in this paper.

References

1. Osband I, Wen Z, Asghari M et al. (2021) Epistemic neural networks. arXiv preprint arXiv:2107.08924
2. Martin F, Ritzhaupt A, Kumar S et al (2019) Award-winning faculty online teaching practices: course design, assessment and evaluation, and facilitation. Internet High Educ 42:34–43
3. Ibrahim AF, Attia AS, Asma’M B et al (2021) Evaluation of the online teaching of architectural design and basic design courses case study: college of architecture at JUST, Jordan. Ain Shams Eng J 12(2):2345–2353
4. Jones RM (2021) Online teaching of forensic medicine and pathology during the COVID-19 pandemic: a course evaluation. J Forensic Leg Med 83:102229
5. Lin H, You J, Xu T (2021) Evaluation of online teaching quality: an extended linguistic MAGDM framework based on risk preferences and unknown weight information. Symmetry 13(2):192
6. Li M, Su Y (2020) Evaluation of online teaching quality of basic education based on artificial intelligence. Int J Emerg Technol Learn (iJET) 15(16):147–161
7. Hou J (2021) Online teaching quality evaluation model based on support vector machine and decision tree. J Intell Fuzzy Syst 40(2):2193–2203
8. Jiang L, Wang X (2020) Optimization of online teaching quality evaluation model based on hierarchical PSO-BP neural network. Complexity. https://doi.org/10.1155/2020/6647683
9. Lv J, Wang X, Huang M (2017) ACO-inspired ICN routing mechanism with mobility support. Appl Soft Comput 58:427–440
10. Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31(3):685–695
11. Burkov A (2019) The hundred-page machine learning book. Andriy Burkov, Quebec City
12. Sun S, Cao Z, Zhu H et al (2019) A survey of optimization methods from a machine learning perspective. IEEE Trans Cybern 50(8):3668–3681
13. Haihong C, Biao H, Feng L, Wengu C (2017) Machine learning principles and applications. University of Electronic Science and Technology Press, Chengdu, pp 2–19
14. Anderson JA (1995) An introduction to neural networks. MIT Press, Cambridge
15. He M (2015) Fundamentals of college computer. Southeast University Press, Nanjing
16. Greydanus S, Dzamba M, Yosinski J (2019) Hamiltonian neural networks. Adv Neural Inf Process Syst 32:1–11
17. What is neural network? https://www.ibm.com/cn-zh/cloud/learn/neural-networks. (2020)
18. Kiranyaz S, Avci O, Abdeljaber O et al (2021) 1D convolutional neural networks and applications: a survey. Mech Syst Signal Process 151:107398
19. Sherstinsky A (2020) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys D 404:132306

20. Jospin LV, Laga H, Boussaid F et al (2022) Hands-on Bayesian neural networks—a tutorial for deep learning users. IEEE Comput Intell Mag 17(2):29–48

21. Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in neural information processing systems, Long Beach, CA, USA, pp. 6405–6416

22. Gal Y, Ghahramani Z (2016) Dropout as a Bayesian approximation: representing model uncertainty in deep learning. International conference on machine learning 48:1050–1059

23. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. In: International conference on machine learning, pp. 8748–8763. PMLR

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.