From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review

Ravil I. Mukhamediev, Adilkhan Symagulov, Yan Kuchin, Kirill Yakunin, and Marina Yelis

Abstract: There are promising prospects on the way to widespread use of AI, as well as problems that need to be overcome to adapt AI&ML technologies in industries. The paper systematizes the AI sections and calculates the dynamics of changes in the number of scientific articles in machine learning sections according to Google Scholar. The method of data acquisition and calculation of dynamic indicators of changes in publication activity is described: growth rate (D1) and acceleration of growth (D2) of scientific publications. Analysis of publication activity, in particular, showed a high interest in modern transformer models, the development of datasets for some industries, and a sharp increase in interest in methods of explainable machine learning. Relatively small research domains are receiving increasing attention, as evidenced by the negative correlation between the number of articles and D1 and D2 scores. The results show that, despite the limitations of the method, it is possible to (1) identify fast-growing areas of research regardless of the number of articles, and (2) predict publication activity in the short term with satisfactory accuracy for practice (the average prediction error for the year ahead is 6%, with a standard deviation of 7%). This paper presents results for more than 400 search queries related to classified research areas and the application of machine learning models to industries. The proposed method evaluates the dynamics of growth and the decline of scientific domains associated with certain key terms. It does not require access to large bibliometric archives and allows to relatively quickly obtain quantitative estimates of dynamic indicators.

Keywords: artificial intelligence; machine learning; deep learning; explainable machine learning; transfer learning; convolution neural networks; recurrent neural networks; transformers; scientometrics; bibliometric indicators

1. Introduction

For successful economically justified development of traditional and new industries, increasing production volumes and labor productivity, we need new technologies related not only to extraction, processing, and production technologies, but also to the collection, processing, and analysis of data accompanying these processes. Of course, one of the most promising tools in this area of development is artificial intelligence (AI). AI already brings significant economic benefits in healthcare [1], commerce, transportation, logistics, automated manufacturing, banking, etc. [2]. Many countries are working out or have adopted their strategies for the use and development of AI [3]. At the same time, there are promising prospects and some obstacles on the way to the widespread use of AI, the overcoming of which means a new round of technological development of AI and expansion of its application sphere.
The evolution of each scientific direction, including AI, is accompanied by an increase or decrease of the interest of researchers, which is reflected in the change of bibliometric indicators. The latter includes the number of publications, the citation index, the number of co-authors, the Hirsch index, and others. The identification of “hot” areas in which these indicators are more important allows us to better understand the situation in science and, if possible, to concentrate the efforts on breakthrough areas.

The field of machine learning (ML) and AI is characterized by a wide range of methods and tasks, some of which already have acceptable solutions implemented in the form of software, while others require intensive research.

In this regard, it would be interesting to consider how the interest of researchers has changed over time and, if possible, to identify those areas of research that are currently receiving increased attention and to focus on them. At the same time, the number of publications in many areas of AI is growing. Therefore, a simple statement of the increase in the number of publications is not enough. In this connection, bibliometric indicators (BI), such as the number of publications, the citation index, the number of co-authors, etc., are widely used to assess the productivity of scientists [4,5]. BI has also been applied to the evaluation of universities [6] and research domains [7]. BI is used for the assessment of policy making in the field of scientific research [8] and the impact of publication databases [9]. The authors of [10] use bibliometric data to build prediction models based on bibliometric indicators and models of system dynamics. In papers [11,12], the mentioned approach is combined with patent analysis. The prediction task is important in the situation of quick technological changes. To do this, the changes in the Hirsch index in time are examined [13] and the concept of a dynamic Hirsch index is suggested [14].

At the same time, bibliometric methods have significant limitations. In particular, numerical indexes are non-linearly dependent on the size of the country and organization [15]. The use of indicators without a clear understanding of the subject area leads to the “quick and dirty” effects [16]. Generally, the BI is a static assessment. To assess the development of scientific fields, it is necessary to consider changes in bibliometric indicators.

In order to identify the logic of changes in publication activity, the differential indicators are implemented in [17]. Their application allows to estimate the speed and acceleration of changes in bibliometric indicators. The implemented indicators can thus more obviously show the growth or decline of the researchers’ interest in certain sections of the AI&ML, characterized by certain key words. In our opinion, the use of dynamic indicators along with full-text analysis allow us to more accurately assess the potentials of research areas.

In this paper, the number of article publications with selected key terms are considered as analyzed indicators. The differential metrics allow to evaluate the dynamics of changes in the usage of selected key terms by the authors of scientific publications, which indirectly indicates the growth or decrease of the interest of researchers in the scientific field designated by this term.

A significant problem in conducting this kind of research is the analysis of the scientific field to identify the key terms that characterize the directions of research and applications. For this purpose, we have made a brief review and systematization of scientific directions included in ML. We have also attempted to assess the applicability of deep learning technologies in various industries. The interpretation of the obtained results, of course, largely depends on the informal analysis performed.

The objectives of the study are as follows:

1. Systematization of AI&ML sections according to literature data.
2. Development of methods for collecting data from open sources and assessing changes in publication activity using differential indicators.
3. Assessment of changes in publication activity in AI&ML using differential indicators to identify fast-growing and “fading” research domains.
The remainder of this work consists of the following sections. The Section 2 provides a brief literature review of AI&ML domains and some classification of fields of study is formed.

In Section 3, we describe the method for analyzing publication activity.

In Section 4, we present the results of the analysis of publication activity based on the previously introduced classification of research areas and problems preventing the successful adaptation of machine learning technologies in production.

Section 5 is devoted to a discussion of the results.

Finally, we summarize the discussion and describe the limitations of our method.

2. Literature Review

Artificial intelligence (AI) is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings [18]. In other words, AI is any software and hardware method that mimics human behavior and thinking. AI includes machine learning, natural language processing (NLP), text and speech synthesis, computer vision, robotics, planning, and expert systems [19]. A schematic representation of the components of AI is shown in Figure 1.

Machine learning significantly realizes the potential inherent in the idea of AI. The main expectation associated with ML is the realization of flexible, adaptive, “teachable” algorithms or computational methods. As a result, new functions of systems and programs are provided. According to the definitions given in [20]:

- Machine learning (ML) is a subset of artificial intelligence techniques that allow computer systems to learn from previous experience (i.e., from observations of data) and improve their behavior to perform a particular task. ML methods include support vector methods (SVMs), decision trees, Bayesian learning, k-means clustering, association rule learning, regression, neural networks, and more.
Neural networks (NN) or artificial NNs are a subset of ML methods with some indirect relationship to biological neural networks. They are usually described as a set of connected elements called artificial neurons, in organized layers.

Deep learning (DL) is a subset of NN that provides computation for multilayer NN. Typical DL architectures are deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), generating adversarial networks (GAN), and more.

Today, machine learning is successfully used to solve problems in medicine [21,22], biology [23], robotics, urban agriculture [24] and industry [25,26], agriculture [27], modeling environmental [28] and geo-environmental processes [29], creating a new type of communication system [30], astronomy [31], petrographic research [32,33], geological exploration [34], natural language processing [35,36], etc.

The classification of machine learning models was considered in [37–40]. ML methods are divided into five classes [41] (Figure 2):

- Unsupervised learning (UL) [42] or cluster analysis.
- Supervised learning (SL) [43].
- Semi-supervised learning.
- Reinforcement learning.
- Deep learning.

Machine learning methods solve problems of regression, classification, clustering, and data dimensionality reduction.

The tasks of clustering and dimensionality reduction are solved by UL methods, when the set of unlabeled objects is divided into groups by an automatic procedure, based on the properties of these objects [44,45]. Among the clustering methods, we can...
list k-means [46], isometric mapping (ISOMAP) [47], local linear embedding (LLE) [48], t-distributed stochastic neighbor embedding (t-SNE) [49], kernel principal component analysis (KPCA) [50], and multidimensional scaling (MDS) [51].

These methods make it possible to identify hidden patterns in the data, anomalies, and imbalances. Ultimately, however, the tuning of these algorithms still requires expert judgment.

SL solves the problem of classification or regression. A classification problem arises when finite groups of some designated objects are singled out in a potentially infinite set of objects. Usually, the formation of groups is performed by an expert. The classification algorithm, using this initial classification as a pattern, must assign the following unmarked objects to this or that group, based on the properties of these objects.

Such classification methods include:
- k-Nearest-Neighbor (k-NN) [52–54].
- Logistic regression.
- Decision tree (DT).
- Support vector classifier (SVM) [55].
- Feed forward artificial neural networks (ANN) [56–58].
- Compositions of algorithms (in particular, boosting [59]).
- Separate group of deep learning networks (DLN), such as long-term memory (LSTM) [60], etc.

Deep learning is a term that combines methods based on the use of deep neural networks. An artificial neural network with more than one hidden layer is considered to be a deep neural network. A network with less than two hidden layers is considered a shallow neural network.

The advantage of deep neural networks is evident when processing large amounts of data. The quality of traditional algorithms, reaching a certain limit, no longer increases with the amount of available data. At the same time, deep neural networks can extract the features that provide the solution to the problem, so that the more data, the more subtle dependencies can be used by the neural network to improve the quality of the solution (Figure 3 [61]).

![Figure 3. Changes in the quality of problem-solving depending on the amount of available data for machine learning algorithms of varying complexity.](image-url)
The use of deep neural networks provides a transition to End-to-End problem-solving. End-to-End means that the researcher pays much less attention to the extraction of features or properties in the input data, for example, extraction of invariant facial features, when recognizing faces, or extraction of individual phonemes in speech recognition, etc. Instead, it simply feeds a vector of input parameters, such as an image vector, to the input of the network, and expects the intended classification result on the output. In practice, this means that, by selecting a suitable network architecture, the researcher allows the network itself to extract those features from the input data that provide the best solution to, for example, the classification problem. The more data, the more accurate the network will be. This phenomenon of deep neural networks predetermined their success in solving the problems of classification and regression.

The variety of neural network architectures can be reduced to four basic architectures (Figure 4):

1. Standard feed forward neural network—NN.
2. Recurrent neural network—RNN.
3. Convolution neural network—CNN.
4. Hybrid architectures that include elements of 1, 2, and 3 basic architectures, such as Siamese networks and transformers.

A recurrent neural network (RNN) changes its state in discrete time so that the tensor $a_{t-1}$, which describes its internal state at time $t-1$, is “combined” with the input signal $x^{(t-1)}$, which comes to the network input at that time and the network generates the output signal $y^{(t-1)}$. The internal state of the network changes to $a_t$. At the next moment, the network receives a new input vector $x^{(t)}$, generates an output vector $y^{(t)}$, and changes its state to $a_{t+1}$, and so on. As input vectors, $x$ can be, for example, vectors of natural language words, and the output vector can correspond to the translated word.
Recurrent neural networks are used in complex classification problems when the result depends on the sequence of input signals or data, and the length of such a sequence is generally not fixed. The data and signals received at previous processing steps are stored in one or another form in the internal state of the network, which allows taking their influence into account in the general result. Examples of tasks with such sequences are machine translation [62,63], when a translated word may depend on the context, i.e., previous or next words of the text:

- Speech recognition [64,65], where the values of the phonemes depend on their combination.
- DNA analysis [66], in which the nucleotide sequence determines the meaning of the gene.
- Classifications of the emotional coloring of the text or tone (sentiment analysis [67]). The tone of the text is determined not only by specific words, but also by their combinations.
- Name entity recognition [68], that is, proper names, days of the week and months, locations, dates, etc.

Another example of the application of recurrent networks are tasks where a relatively small sequence of input data causes the generation of long sequences of data or signals, for example:

- Music generation [69], when the generated musical work can only be specified in terms of style.
- Text generation [70], etc.

Convolutional neural network: In the early days of the computer vision development, researchers made efforts to teach the computer to highlight characteristic areas of an image. Kalman, Sobel, Laplace, and other filters were widely used. Manual adjustment of the algorithm for the extraction of the characteristic properties of images allowed to achieve good results in particular cases, for example, when the images of faces were standardized in size and quality of photographs. However, when the foreshortening, illumination, and scale of images were changed, the quality of recognition deteriorated sharply. Convolutional neural networks have largely overcome this problem.

The convolutional neural network shown in Figure 4 has as input an image matrix, in which data passes sequentially through the three convolutional layers: Conv1 of dimension (2, 3, 3), Conv2 of dimension (5, 2, 4), and Conv3 of dimension (1, 3). The application of convolutional nets allows distinguishing complex regularities in the presented data invariant to their location in the input vector. For example, to select vertical or horizontal lines, points, and more complex figures and objects (eyes, nose, etc.).

Image processing includes problems of identification (cv1), verification (cv2), recognition (cv3), and determination (cv4) of visible object characteristics (speed, size, distance, etc.). The most successful algorithm for cv1 and cv3 problems is the YOLO algorithm [71,72], which uses a convolutional network to identify object boundaries “in one pass”.

Hybrid architectures: The cv2 problem is often solved using Siamese networks [73] (Figure 4), where two images are processed by two identical pre-trained networks. The obtained results (image vectors) are compared using a triplet loss function, which can be implemented as a triplet distance embedding [74] or a triplet probabilistic embedding [75].

The network is trained using triples \((x_a, x_p, x_n)\), where \(x_a\) (“anchor”) and \(x_p\) (positive) belong to one object, and \(x_n\) (negative) belongs to another. For all three vectors, the embeddings \(f(x_a)\), \(f(x_p)\), and \(f(x_n)\) are calculated. The threshold value of alpha \((\alpha)\) is set beforehand. The network loss function is as follows:

\[
L(x_a, x_p, x_n, f) = \frac{1}{N} \sum_{i=1}^{N} \left[ f(x_a^i) - f(x_p^i) \right]_2^2 - \left[ f(x_a^i) - f(x_n^i) \right]_2^2, \quad (1)
\]

where \(N\) is the number of objects.

The triplet loss function “increases” the distance between embeddings of images of different objects and decreases the distance between different embeddings of the same object.
BERT (bidirectional encoder representations from transformers) [76], ELMO [77], GPT (generative pre-trained transformer), and generative adversarial networks [78] have recently gained great popularity and are effectively used in natural language processing tasks.

Obstacles to the application of AI&ML: There are promising prospects for the widespread use of AI, as well as a number of problems, the overcoming of which means new opportunities for adapting AI technologies in production and a new round of technological development of AI.

The scientific community distinguishes social (fear of AI), human resources (shortage of data scientists) and legal [79], organizational and financial [80], as well as a number of technological problems associated with the current level of AI&ML technology development. They include [81]: data problems (data quality and large volume of data), slow learning process, explaining the results of ML models, and significant computational costs.

The literature survey helps to identify a set of key terms for quantitative analysis of publication activity. However, this approach retains a certain amount of subjectivity in the selection of publications.

In addition to the systematization of AI&ML sections, their evolution is also of interest. Let us assess the dynamics of changes in the number of scientific publications aimed at the development of individual scientific domains and overcoming the aforementioned technological limitations of AI&ML.

3. Method

Publication activity demonstrates the interest of researchers in scientific sections, which are briefly described by some sets of terms. Obviously, new and promising in the eyes of the scientific community, thematic sections are characterized by an increased publication activity. To identify such sections and their comparative evaluation in the field of AI and ML learning, we will use the method described in [17]. The paper proposes dynamic indicators that allow to numerically estimate the growth rate of the number of articles and acceleration. The indicators allow us to estimate the scientific field without regard to its volume, which is important for new fast-growing domains that do not yet have a large volume of publications.

The dynamic indicators (D1—speed and D2—acceleration) of the j-th bibliometric indicator \( s_j^{(db,k)} \) at time \( t_n \) can be calculated as follows:

\[
D1_j^{(db,k)}(t_n) = w_1 j \times \frac{ds_j^{(db,k)}(t_n)}{dt}, \tag{2}
\]

\[
D2_j^{(db,k)}(t_n) = w_2 j \times \frac{d(s_j^{(db,k)}(t_n)/dt)}{dt}, \tag{3}
\]

where \( k \) is the search term in database \( db \), and \( w_1 \) and \( w_2 \) are empirical coefficients that regulate the “weight” of the \( s_j^{(db,k)} \).

In our case, \( s_j^{(db,k)}(t_n) \) is the number of articles in \( t_n \)—the year selected using the search query \( k \) in the Google Scholar database. Weights \( w_1 \) and \( w_2 \) are taken as 1. For example, the search query \( k = \text{“Deep + Learning + Bidirectional + Encoder + Representations + from + Transformers”} \) provided the following annual publication volumes: 13, 692, and 1970 in 2018, 2019, and 2020, respectively. Bibliometric databases often provide an estimate of the number of publications at the end of the year. Approximation of the obtained numerical series is performed with the help of the polynomial regression model, in which, as we know, regression coefficients are calculated for the function of the hypothesis of the form:

\[
h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + ... + \theta_n x^n, \tag{4}
\]
where \( \theta_i \in \Theta \) are regression parameters, and \( n \) is order or degree of regression dependence. Assessment of the quality of the constructed regression dependence is performed, as a rule, using the coefficient of determination:

\[
R^2 = 1 - \frac{SS_{res}}{SS_{tot}}
\]  

(5)

\[
SS_{res} = \sum_{i=1}^{m_k} (y^{(i)} - h^{(i)})^2
\]  

(6)

\[
SS_{tot} = \sum_{i=1}^{m_k} (y^{(i)} - \bar{y})^2, \bar{y} = \frac{1}{m_k} \sum_{i=1}^{m_k} y^{(i)},
\]  

(7)

where

- \( y^{(i)} \) — actual value,
- \( h^{(i)} \) — calculated value (hypothesis function value) for the \( i \)-th example,
- \( m_k \in m \) — part of the training sample (sets of marked objects).

In existing libraries, \( R^2 \) is denoted by \( r_2 \_score \). The best value of \( r_2 \_score = 1 \).

Increasing the order of the regression \( n \) allows us to obtain a high value of the coefficient of determination, but usually leads to overtraining of the model. In order to avoid overtraining and to ensure a sufficient degree of generalization, the following rules of thumb are used:

- For each search query, the regression order is chosen individually, starting from \( n = 3 \) to ensure \( r_2 \_score \geq 0.7 \). As soon as the specified boundary is reached, \( n \) is fixed and the selection process stops.
- Since we are most interested in the values of dynamic indicators for the last year, we used the last value (number of publications for 2020) and the value equal to half of the growth of articles achieved at the end of 2020, which we conventionally associate with the middle of the year, as a test set on which \( r_2 \_score \) is determined.

Data processing complex for calculation of D1 and D2 indicators includes scraper, preprocessing, and regression calculation with selection of regression order, providing \( r2 \geq 0.7 \) on the specially formed test set and calculation of D1 and D2 indicators (Figure 5).

![Figure 5. Scheme of data collection and processing.](image)

The scraper uses the requests library to retrieve the HTML page of each search query and then uses the BeautifulSoup library to select the necessary information from it. Since scholar.google.com (accessed on 8 June 2021) has protection against robots, the app.proxiesapi.com (accessed on 8 June 2021) service was used to provide a proxy server. Thanks to this, it was possible to avoid captcha.

The matching of the input data to the query is determined by the capabilities of the Google search engine. Since it is impossible to carry out a full analysis of all articles, we...
have performed selective validation. For this, requests were made for articles of 2020. Then, by manual analysis of the text of the articles, the identification of the correspondence of the 10th and 100th articles to the request was performed. The results were assessed on a three-point scale: 2—full compliance, 1—partial compliance, 0—does not correspond to the semantics of the query. The verification results are provided in Appendix A. They show that only in 3 cases out of 48 did the selected articles not fully correspond to the request.

Pre-processing consists in the formation of a data-frame containing only the necessary information (search query and the numeric series of the annual number of publications). No additional processing is required.

The annual number of publications is approximated using a polynomial regression model. The result of each query is time series. The time series is approximated as described above. The obtained regression model is used to calculate the D1 and D2 indicators and make predictions.

The indicators calculated for the last year reflect the dynamics of changes in the interest of the scientific community in the relevant scientific sections at the time of the study.

Data collection and processing were realized using python, numpy, sklearn, and pandas.

4. Results and Discussion

The total number of AI&ML publications and changes in some of the declining and growing areas from the 2015–2020 years are shown in Figure 6. The data from the table in Appendix B were used to create the figure. It can be seen that the total number of articles, which peaked at 674,000 in 2019, has significantly decreased in 2020 (594,000). The graph shows a slowdown in overall growth since 2018. The reason for this is the decrease since 2017 in the number of articles with the keywords Robotics, Expert Systems, etc., which could not be compensated by the growth of deep learning sections. Calculation results and software are provided as supplementary materials.

![Figure 6. The annual number of publications in the fields of AI&ML.](image)

The described set of article counts and calculation of D1 and D2 indicators has been applied to assess the dynamic indicators of the main sections of AI (Figure 7), the popularity of the main models of machine learning and deep learning (Figures 8 and 9), and publication activity related to explanatory AI applications and modern ML models in economics (Figures 10 and 11).
Figure 7. The popularity of AI sections. In gray is the number of articles on a logarithmic scale. In turquoise is D1, in orange is D2.

Figure 8. Using popular models of unsupervised learning (UL), machine learning (ML), and deep learning (DL). In gray is the number of articles on a logarithmic scale. In turquoise is D1, in orange is D2.
Figure 9. Researcher interest in explanatory machine learning methods and the development of datasets and corpora of texts. In gray is the number of articles on a logarithmic scale. In turquoise is D1, in orange is D2.

Figure 10. Application of EML and EAI to industries. In gray is the number of articles on a logarithmic scale. In turquoise is D1, in orange is D2. The full version is provided in Appendix C.
Figure 11. Applying modern deep learning models to industries. The full version is provided in Appendix D.

Since some domains do not have a large number of publications, the graphs show the maximum number of articles (mnp) in the domain during the analysis period on a logarithmic scale—$\log_{10}(mnp)$.

The interpretation of the indicator values is as follows:

- When both values (D1, D2) are positive, it indicates an accelerated growth of the number of articles in this domain.
- When D1 is negative and D2 is positive, this indicates a slowdown in the number of items.
- When D1 is positive and D2 is negative, it indicates a slowdown in the growth of the number of items.
- When both values are negative, it indicates an accelerated decrease in the number of items.

Figure 8 shows a significant increase in publication activity in the field of deep learning, which was expected. The maturity and relevance of the scientific field can be assessed by the number of review articles. We performed this analysis for the machine learning sections (Figure 2) by counting the articles that contain the terms review/overview/survey in the title. The largest number of such reviews is in the deep learning domain (852 for 2020, an increase of more than 20 times since 2015), reinforcement learning (101 articles for 2020, an increase of more than 10 times since 2015), and supervised learning (38 articles in 2020, where the increase since 2015 was more than seven-fold).

Figure 8 shows a significant and accelerating increase in publication activity in the domain of explainable machine learning and transformers’ applications. Explaining the results of machine learning models is a serious problem, preventing widespread use of AI in healthcare [82], banking, and many other fields [83].
The easiest way to interpret a linear regression model is to use coefficients $\theta$ to determine the weights:

$$h_\theta = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n,$$

where $\theta_i \in \Theta$, $x_i \in X$, $h_\theta$—linear regression model hypothesis function.

A complex machine learning model is a “black” box, hiding the mechanism for obtaining results. To turn it into a “white” or “gray” box, methods are used to estimate the influence of input parameters on the final result. There are basic methods Treeinterpreter, DeepLIFT, etc., however recently, local interpretable model-agnostic (LIME) [84] and SHapley Additive exPlanations (SHAP) [85] have become very popular. LIME creates an interpretable model, for example, a linear one, that learns on small perturbations of the parameters of the object being evaluated (“noise” is added), achieving a good approximation of the original model in this small range. However, for complex models with a significant correlation of properties, linear approximations may not be sufficient.

SHAP is designed to work when there is a significant relationship between features. In general, the method requires over-training of the model on all subsets, $S \subseteq n$, where $n$ is the set of all features. The method assigns a value of importance to each property, which reflects the effect on the model prediction when this property is enabled. To calculate this effect, the model $f(S \cup \{i\})$ is trained with this property and the other model $f(S)$ is trained with the excluded property. Then, the predictions of these two models are compared at the current input signal $f(S \cup \{i\} (xS \cup \{i\})) - f(S)$, where $xS$ represents the values of the input properties in the set $S$. Since the effect of eliminating a feature depends on other features in the model, this difference is calculated for all possible subsets, $S \subseteq n \setminus \{i\}$. Then, the weighted average of all possible differences is calculated:

$$\varphi_i = \sum_{S \subseteq \{1,2,...,n\} \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} (f(S \cup \{i\}) - f(S))$$

This is the assessment of the importance (influence) of the properties (features) on the assessment of the model. This approach, based on game theory, according to the conclusions of the authors of the algorithm, provides a common interpretation and suitability for a wide range of machine learning methods. Although the method is used in decision support systems [86], however, interpretation of the influence of individual model parameters is possible if they have a clear meaning.

Figures 10 and 11 show the applicability of machine learning models to industries. There has been a significant increase in publication activity related to the terms: “dataset”, “precision agriculture”, “precision farming”, etc. In the field of deep learning, the solution of many problems depends on the volume and quality of datasets such as ImageNet [87], Open Images [88], COCO Dataset [89], FaceNet. However, they may not be sufficient for specific tasks. The problem of data scarcity in computer vision is overcome with the use of synthetic sets created with 3D graphics editors [90], game engines, and environments [91–94]. Such DSs, in particular, have been used to train unmanned vehicles. Synthetic datasets are used to train unmanned vehicles [95] and in other fields [96]. Recently, generative adversarial networks [97,98] have also been used for their generation.

The regularities of publication activity are such that some terms (robotics, supervised learning, machine vision, regression, etc.) are used less and less frequently. In the field of NLP, significant growth is observed in the domains of topic modeling, text generation, and question answering. At the same time, according to the available data, the growth in the domain of sentiment analysis is slowing down. As for the domain expert system, it is characterized by a slowdown in the decrease in the number of publications, unlike the domain recurrent neural network, which, while showing an observed increase in the number of articles, is nevertheless characterized by a significant slowdown of this increase. It can be assumed that recurrent network research is shifting towards new architectures of neural networks, explanation of results, etc. For example, we can see a pronounced growth
of publication activity related to the terms Siamese neural networks and convolutional neural networks.

Scientific articles are aimed at overcoming the limitations of AI&ML technology. However, they make it possible to assess which noun groups of algorithms are most in demand in practice. In particular, in [99], it was revealed that deep learning technologies demonstrate a high increase in publication activity as applied to healthcare. Current analysis confirms that healthcare is one of the popular application domains for deep learning. At the same time, models of transformers (BERT) and generative adversarial networks show very high rates of D1 and D2 in combination with the term healthcare. BERT, as applied to many industries (development, production, manufacturing, communication, electricity, supply chain, etc.), demonstrates a high rate of publication growth. At the same time, the selection for the search terms Electricity Deep + Learning and Social + Services Deep + Learning shows a sharp slowdown in the growth of publications. This phenomenon can be interpreted as a possible shift in researchers’ interest in new terms describing modern deep learning models.

The negative correlation between the number of articles and the indicators D1, D2, and r2_score (Figure 12) allows us to conclude that the domains with a large number of publications are characterized by a decrease in the dynamics of publication activity and model error.

![Feature Correlation Heatmap](image)

**Figure 12.** Correlation heatmap of bibliometric indicators.

Despite the simplicity of the model, it can also be used for prediction. For example, predicting the number of articles one year ahead has an average error of about 6% with a standard deviation of 7%.

The forecast accuracy naturally decreases with an increase in the period. Depending on the queries, the mean squared error is from 0.017 to 0.13 and 0.04 to 0.19 for a 2- and 3-year forecast, respectively. However, the maximum value of mean squared error rises to 0.82, which significantly reduces the value of the prediction.

However, based on the values of D1 and D2, the following assumptions can be made:

- In general, the number of publications in the AI&ML domain will decrease.
- New domains such as applications of transformers and explainable machine learning will see rapid growth.
• Classic machine learning models such as SVM, k-NN, and logistic regression will attract less attention from researchers.
• The number of articles on clustering models will continue to increase.

5. Conclusions

The world of AI is big. The fastest-growing area of research is machine learning, and within it, deep learning models. New results, as well as applications of previously proposed networks, appear almost daily. This area of research and applications includes a large family of networks for text, speech, and handwriting recognition, networks for image transformation and stylization, and networks for processing temporal sequences. Siamese networks are a relatively new direction of applications of deep neural networks, which show high results in recognition tasks, networks for object identification that provide confident object identification, transformer models that solve the problem of recognition and text generation, etc.

Research Contribution.

In this paper, we systematized the sections of AI and evaluated the dynamics of changes in the number of scientific articles in the machine learning domains according to Google Scholar. The results show that, firstly, it is possible to identify fast-growing and “fading” research domains for any “reasonable” number of articles (>100), and secondly, the prediction of publication activity is possible in the short term, with sufficient accuracy for practice.

Research Limitations.

The method we used has some limitations, in particular:

1. For all the depth of the informal analysis, the set of terms is still set by the researcher. Consequently, some of the articles that are part of the section under study may be left out, and, conversely, some publications may be incorrectly attributed to the topic in question. We also cannot guarantee the exhaustive completeness and consistency of the empirical review performed.

2. This analysis does not take into account the fact that the importance of a particular scientific topic is determined not only by the number of articles, but also by the volume of citations, the “weight” of the individual characteristics of the authors, the quality of the journals, and so on.

3. The method does not evaluate term change processes and semantic proximity of scientific domains.

Research Implications.

The obtained estimates, despite some limitations of the applied approach, correspond to empirical observations related to the growth of applications of deep learning models, the construction of explicable artificial intelligence systems, and the increase in the number and variety of datasets for many machine learning applications. The analysis shows that the efforts of the scientific community are aimed at overcoming the technological limitations of ML. In particular, methods for generating datasets, explaining the results of machine learning systems, and accelerating learning have already been developed and successfully applied in some cases. However, new solutions are needed to overcome the described limitations for most AI applications. As soon as this happens, we will witness a new stage in the development of AI applications.

Future Research.

In our opinion, overcoming the disadvantages of the described method of evaluating publication activity lies in the use of advantages of topic modeling and analysis of text embeddings. We plan to use classical topic modeling [100] and the embedded topic model [101] for automatic clustering of the corpus of scientific publications and identifying related terms. In addition, a more advanced scraper will provide data on the number of citations of articles and the quality of scientific journals.
The program and results of current calculations can be downloaded at https://www.dropbox.com/sh/fkfw3a1hkf0suvcc/AACRZ7v9qympen_ht00jejF6a?dl=0 (accessed on 8 June 2021).

**Supplementary Materials:** The program and results of current calculations can be downloaded at https://www.dropbox.com/sh/fkfw3a1hkf0suvcc/AACRZ7v9qympen_ht00jejF6a?dl=0 (accessed on 8 June 2021), figures and appendices are available online at https://www.dropbox.com/sh/eothtmbq1idlunn/AAAHdhTBkpQ98bs6jsUf1Cp5a?dl=0 (accessed on 8 June 2021).

**Author Contributions:** Conceptualization, R.I.M. and A.S.; methodology, R.I.M.; software, R.I.M., A.S. and K.Y.; validation, Y.K., K.Y. and M.Y.; formal analysis, R.I.M.; investigation, A.S.; resources, R.I.M.; data curation, A.S. and K.Y.; writing—original draft preparation, R.I.M.; writing—review and editing, Y.K. and M.Y.; visualization, R.I.M. and K.Y.; supervision, R.I.M.; project administration, Y.K.; funding acquisition, R.I.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan, grant number AP08856412.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available in a publicly accessible repository that does not issue DOIs. The program and results of current calculations can be downloaded at https://www.dropbox.com/sh/fkfw3a1hkf0suvcc/AACRZ7v9qympen_ht00jejF6a?dl=0 (accessed on 8 June 2021), figures and appendices are available online at https://www.dropbox.com/sh/eothtmbq1idlunn/AAAHdhTBkpQ98bs6jsUf1Cp5a?dl=0 (accessed on 8 June 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix A**

| Query                                      | 10th article 2020 | 100th article 2020 | Last article 2020 (if articles count less than 100) | Number of articles |
|--------------------------------------------|-------------------|--------------------|-----------------------------------------------------|-------------------|
| Semi-supervised+Learning Graph+Based+Algorithms | 2                 | 2                  |                                                     | 264               |
| Machine+Learning+Supervised+Learning       | 2                 | 2                  |                                                     | 38400             |
| Machine+Learning+Deep+Learning             | 2                 | 2                  |                                                     | 96300             |
| Deep+Learning+Recurrent+Neural+Networks     | 2                 | 2                  |                                                     | 30500             |
| Deep+Learning+Deep+Feed+Forward+Networks    | 2                 | 2                  |                                                     | 157               |
| Machine+Learning+Dataset                   | 2                 | 2                  |                                                     | 64400             |
| Explainable+Machine+Learning               | 2                 | 2                  |                                                     | 14500             |
| Explainable+Artificial+Intelligence         | 2                 | 2                  |                                                     | 16100             |
| Deep+Learning+Transformers                 | 2                 | 2                  |                                                     | 16800             |
| Natural+Language+Processing+Corpora        | 2                 | 1                  |                                                     | 18700             |
| Machine+Learning+Logistic+Regression       | 2                 | 2                  |                                                     | 31000             |
| Machine+Learning+Decision+Tree             | 2                 | 2                  |                                                     | 39600             |
| Unsupervised+Learning+Dimension+Reduction  | 2                 | 2                  |                                                     | 13000             |
| Deep+Learning+Feed+Forward+Neural+Network  | 2                 | 2                  |                                                     | 7010              |
| Unsupervised+Learning+Principal+Components+Analysis | 2             | 2                  |                                                     | 26400             |
| Manufacturing+Deep+Learning                | 2                 | 2                  |                                                     | 39                |
| Production+Deep+Learning                   | 2                 | 2                  |                                                     | 29800             |
| Production+Deep+Learning+BERT               | 2                 | 2                  |                                                     | 19                |
| Development+Deep+Learning+BERT              | 2                 | 2                  |                                                     | 38                |
| Gas+Deep+Learning+GAN                      | 2                 | 2                  |                                                     | 20                |
| Retail+Explainable+Artificial+Intelligence  | 2                 | 2                  |                                                     | 160               |
| Gas+Dataset                                | 2                 | 1                  |                                                     | 34400             |
| Healthcare+Explainable+Machine+Learning    | 2                 | 2                  |                                                     | 607               |
| Precision+Agriculture                      | 2                 | 1                  |                                                     | 171               |
| Explainable+Machine+Learning               | 2                 | 2                  |                                                     | 24300             |

Figure A1. Validation of input data.
Appendix B

Figure A2. Annual number of publications in the field of AI&ML.
Appendix C

| Industry Sector | Explainable AI&ML |
|-----------------|-------------------|
| Precision-Agriculture | Explainable AI & ML for monitoring |
| Healthcare | Explainable AI & ML for diagnostics |
| Gas Exploration | Explainable AI & ML for reservoir management |
| Precision-Agriculture | Explainable AI & ML for crop yield prediction |
| Electric Power | Explainable AI & ML for grid optimization |
| Precision-Agriculture | Explainable AI & ML for irrigation systems |
| Retail | Explainable AI & ML for customer behavior analysis |
| Agriculture | Explainable AI & ML for crop health monitoring |
| Manufacturing | Explainable AI & ML for production planning |
| Manufacturing | Explainable AI & ML for supply chain optimization |
| Supply Chain Management | Explainable AI & ML for inventory management |
| Energy | Explainable AI & ML for energy efficiency improvements |
| Manufacturing | Explainable AI & ML for defect detection |
| Financial Services | Explainable AI & ML for fraud detection |
| Manufacturing | Explainable AI & ML for quality control |
| Manufacturing | Explainable AI & ML for maintenance planning |
| Manufacturing | Explainable AI & ML for energy management |
| Manufacturing | Explainable AI & ML for resource allocation |

Figure A3. Explainable AI&ML in industry.
Figure A4. Explainable AI&ML in industry. Indicators D1 and D2.
Appendix D

Figure A5. Applying modern deep learning models to industry.
Figure A6. Applying modern deep learning models to industry. Indicators D1 and D2.
100. David, B.M.; Andrew, N.Y.; Michael, J.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

101. Dieng, A.B.; Ruiz, F.J.; Blei, D.M. Topic modeling in embedding spaces. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 439–453. [CrossRef]