Characteristics of Machine Learning Research with Impact

Louis Kirsch
http://louiskirsch.com

University College London

May 2018

1 Introduction

In recent years there has been a huge surge in machine learning publication volume and industry investments [1, 2]. The number of publications grew far beyond what any researcher can read. ArXiv [3], one of the major preprint platforms for the field of machine learning, shows an exponential increase in papers uploaded to the platform as shown in Figure 1. A preprint platform allows papers to be uploaded prior to publication and without any peer-review process. Additionally, the field is becoming very competitive with research institutions and companies such as Google, Facebook, DeepMind, Apple, Microsoft, Uber, and many others involved. New research achievements are quickly outdated and surpassed in a matter of a few months or years. This poses the question, what kind of research outlasts the test of time? What characterizes the most impactful research and how can impact be measured? What kind of research areas and questions should the machine learning community focus on to maximize impact, and what kind of methodologies should be used?

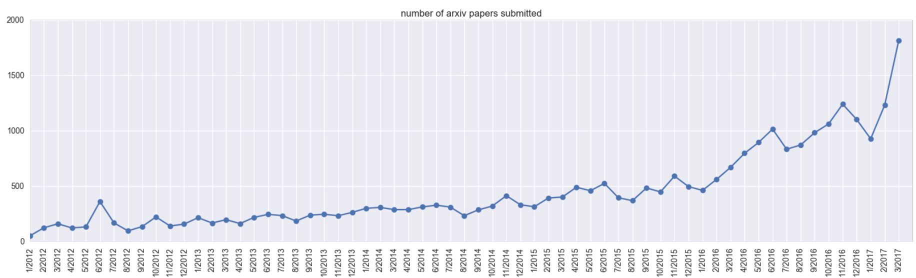


Figure 1: The increase of research paper volume on major preprint platform ArXiv [3]. Reproduced from [4].

2 Impact in science and machine learning research

To identify or carry out impactful research we must first investigate what constitutes impact in machine learning research. In a subject-unspecific context impact of science is measured for the purpose of acceptance in journals, conferences, determining funding, or policy making [5]. With limited available resources, politicians need to decide which research to fund and measure whether these expectations were met. Though, in the field of machine learning, funding is less of an issue due to little material necessities apart from computational resources. Additionally, due to its direct applicability for industry, many companies are interested in funding these efforts for competitive advantages [1].

The oldest method of impact measurement is quality control in the form of peer-review [6]. Acceptance or even rating of research through peer-review constitutes recognition of quality within the research community. Nevertheless, this does not necessarily indicate how long the research will stay relevant or its impact on issues in society. In the context of machine learning Wagstaff [7] claims that most peer-reviewed machine learning contributions only involve a fraction of an impact-generating process. Wagstaff divides the machine learning research process into three stages: The preparation stage includes phrasing a problem as a machine learning task, collecting data, and the selection and generation of features as necessary. Secondly, the machine learning contribution includes choosing or developing the right algorithm, then selecting metrics, and conducting experiments. Thirdly, in the impact stage, results need to be interpreted, published to the relevant user community, and users persuaded to adopt the technique. In his work [7], he criticizes that the machine learning research community mostly only advocates the second stage, not giving any incentives to pursue the first and third stage. As such, his definition of machine learning research should be user impact-oriented and needs to encompass applicability instead of just the development of fundamental learning algorithms. One might counterargue with the fact that machine learning research today has an enormous industry support [1, 2] that deals mostly with stage one and two, while the second stage shall remain the main focus of conferences and journals.

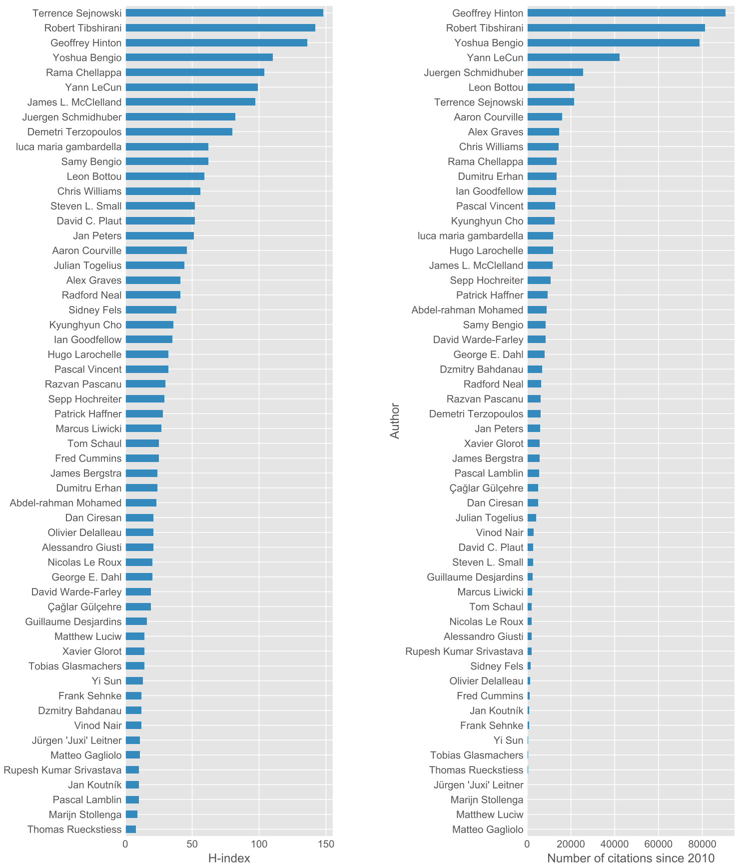
A more quantitative and most universal measurement of impact is citation count [8]. While citation count may measure usefulness to the science community, there is no standard measure for the benefit of research to society [5]. Although, because of the previously mentioned direct applicability of machine learning to industry and many publications directly originating from industry [2], citation count might indeed correlate with benefit to society in the case of machine learning. Impact may also be measured in terms of public appearance through altmetrics [9]. This constitutes web-based measures such as views, downloads, or shares in social media of a research piece or blog post. In particular companies, such as DeepMind, have started advertising their research significantly on platforms such as Twitter.

We hypothesize that this kind of direct measurement on short time scales may be particularly uncorrelated with breakthroughs in machine learning research. As we explore in the following, while direct impact on society is usually driven by applications of machine learning research, the fundamental learning algorithms may take years or decades to benefit society in any measurable manner. In that regard, machine learning research may be similar to research in healthcare that takes very long to progress [5]. We will investigate instances of this phenomenon in the field of deep learning, a subset of machine learning, with regard to contributions such as backpropagation and the long short-term memory units (LSTMs).

Highly cited research in machine learning is often not the original research but books or reviews [10–16]. This is a phenomenon that shows that citation count is not necessarily proportional to the novelty of research or the impact. The reviews and books have to be based on original research that might not be cited as often due to researchers citing aggregations. Jürgen Schmidhuber looked at this issue in the context of one of the most influential algorithms in deep learning called backpropagation [17]. It is a key ingredient for the optimization of most of today's deep learning models and goes back to work in the early 1960s [18–20] but being popularized much later. These original papers, among many others, do not get cited often even though their significance for today's deep learning research is paramount. This conforms with findings of studies by Marx and Bornmann showing that the papers essential for the beginning of a revolution are rarely cited [21, 22]. In fact, the idea of backpropagation is really hard to trace back to a specific researcher. Instead, the algorithm has been repeatedly revised and generalized. Then it took several more decades until the algorithm was rediscovered and used in large-scale applications and state of the art research, for instance in image classification [23]. We can conclude that correct attribution is not always guaranteed and often neglected by researchers, the number of citations is not necessarily proportional to impact, and ideas often need long periods of time until they are recognized as valuable. Even then the original authors might be hard to identify [17].

3 A quantitative analysis

Because there are few quantitative measurements of impact and citation in the field of machine learning we used and modified the open source library Scholarly [24] to query Google Scholar for the most citations in the field of deep learning and artificial intelligence. Just like Google's search engine, Google Scholar aggregates information from web pages. In this case, publications anywhere on the web are located and presented to the researcher, including statistics like authors, publication date, venue, citations, and others. A simple measure of impact could be defined by the search for authors with high citation counts that involve fields of interest 'deep learning' and/or 'artificial intelligence'. Unfortunately, this query results in many highly cited authors with publications in tangent fields such as 'physics', 'big data', 'simulation', 'biology', and others due to the wide range of applications that are based on deep learning. Instead, we focus on actor-network-theory [25] and inspect the co-author network around three deep learning titans [26] 'Geoffrey Hinton', 'Yoshua Bengio', and 'Jürgen Schmidhuber'. Of course, these three researchers are only a tiny selection of many very successful researchers in the field of deep learning. The coauthors



(a) The h-index

(b) Number of citations since 2010

Figure 2: The coauthor network of three deep learning titans 'Geoffrey Hinton', 'Yoshua Bengio', and 'Jürgen Schmidhuber'.

are limited to the 20 most frequent for each author, as given by Google Scholar. Because of outliers that are highly cited we use the popular h-index as a measure of citation. For an h-index of h, any researcher needs to have h papers with at least h citations each. Figure 2a shows the three authors and their coauthors sorted by their h-indices. Furthermore, because deep learning has gained influence in particular from around 2010 [26], we show the total number of citations since 2010 in Figure 2b. From both figures, it is quickly evident that the majority of citations is among very few authors. Also, it is a good method to discover related authors and publications with high impact or high impact potential.

One very famous researcher, Geoffrey Hinton, that coined the term backpropagation with a paper in 1986 [27] worked in the field for several decades until the algorithm showed large-scale success in 2012 [23]. One often advocated reason, why large-scale success was only achieved much later, is the fact that computers got several magnitudes faster, in particular through the use of GPUs instead of

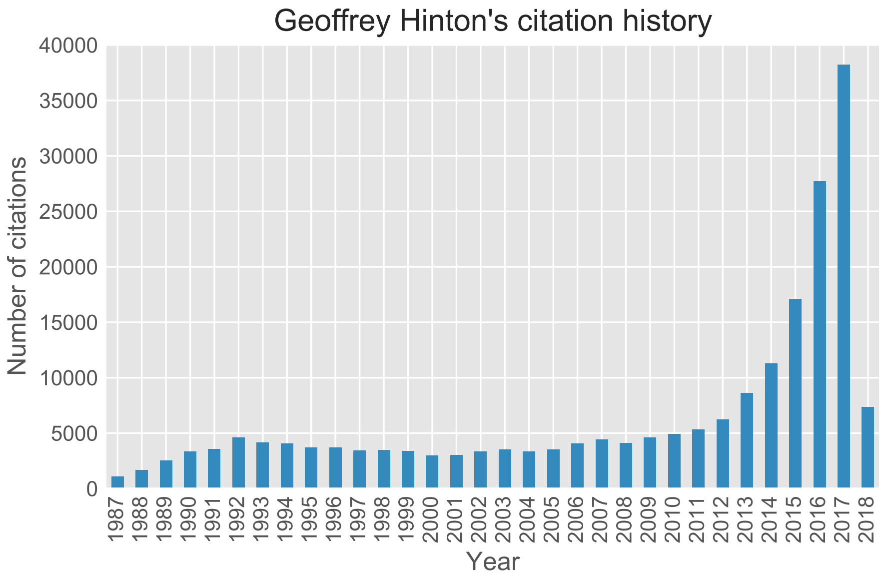


Figure 3: Geoffrey Hinton; a deep learning pioneer's citation history. While the original backpropagation algorithm was invented in 1986 [27] it took until around 2012 that an image classification paper showed the significance of these contributions [23]. Soon after, deep learning started to take off and the decade-long work was cited more often.

CPUs, and datasets got significantly larger [26]. Before 2012, most researchers were highly skeptical of deep learning and focused on other methods in the field of machine learning [28]. This is also reflected in the number of citations that only increased after this practical success in 2012 as shown in Figure 3. This phenomenon has been shown to be generally true in science. The significance of publications is often only discovered decades later [29, 30]. These findings question whether research with true impact in machine learning, such as backpropagation, can be planned. In the case of Geoffrey Hinton and others, researchers trusted their intuition, despite rejection in large parts of the community. While there is certainly no guarantee that this approach is successful it could inspire other machine learning researchers to explore less popular areas that seem promising to them. Another learning might be that research not only needs to be proven scientifically but also showcased on large applications, such as image classification. This not only improves impact in terms of applicability in industry but also recognition and citation in the scientific community.

A similar case can be made for the Long-Short-Term-Memory networks originally proposed in 1997 by Hochreiter and Schmidhuber [31]. Starting with large-scale success in handwriting-recognition in 2007 and 2009 [32, 33] and speech recognition in 2013 [34] LSTMs became widely used both in academia and industry. Similar to Geoffrey Hinton's work this delay is directly reflected in citation counts as shown in Figure 4.

In general, science has many papers which are barely cited and few with high

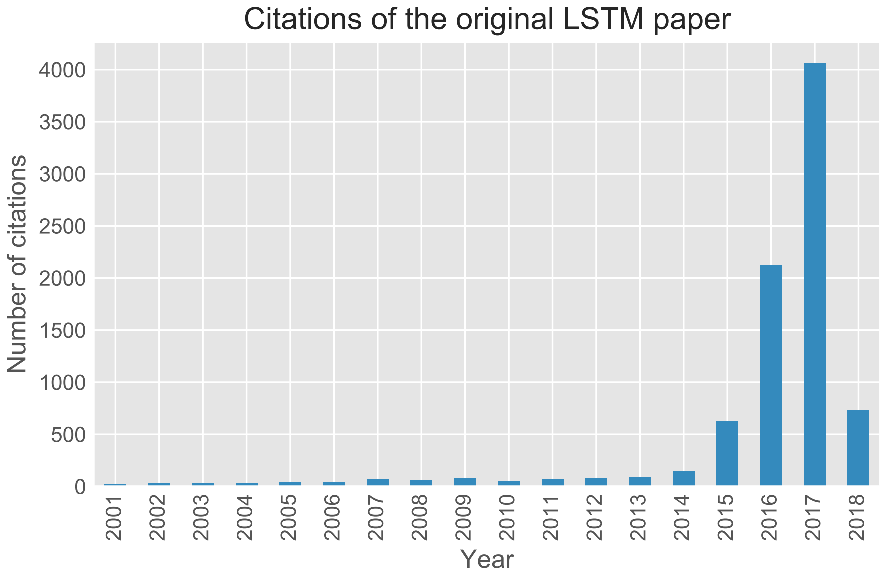


Figure 4: While the LSTM has been published in 1997 [31] it took several years and success in the form of applicability for real-world challenges such as handwriting [32, 33] and speech recognition [34] to be widely recognized and used in other research.

citation count [35]. We further analyze Geoffrey Hinton's publications to test whether this is true for deep learning research. Among the most cited papers rank backpropagation related papers such as [27] with around 42 thousand citations, and the already mentioned image classification paper [23] with 21 thousand citations. Other highly cited work includes learning algorithms for deep belief nets [36] with 7 thousand citations and a review paper on deep learning from only 2015 [14] with 6 thousand citations. Of course, these papers are the exception in the distribution of citations for papers even for famous researchers such as Geoffrey Hinton. The median of citations for all his papers is just at 17 citations. To get an idea how skewed the distribution is, we plot a histogram of the 80% less cited papers in Figure 5. If we had included the entire distribution, the previously mentioned outliers would render the visualization useless. Of this 80% the median is only at 8 citations per paper. Clearly, from Figure 5 it can be observed that most papers are well below 25 citations and a huge amount was not cited at all or only very few times. Together with the previous observation from Figure 2b we can support the findings by Bornmann and Ioannidis that science is determined by few scientists and publications [37, 38] also in the field of deep learning. Popper's formulation of science being defined by 'trial and error' [39] becomes very evident in the field of machine learning as demonstrated by a large number of barely cited publications even by very successful researchers. From the observations in this section, we can also conclude that large numbers of citations indeed are a good indicator of impact, both for applicability in academia, as well as in industry. But as we have seen, this often is only true in the long-term and fewer citations are not necessarily a sign of

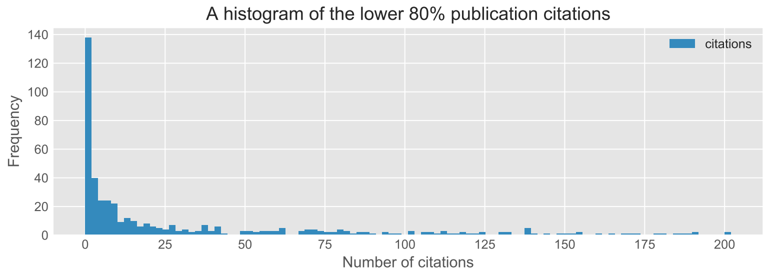


Figure 5: A histogram of Geoffrey Hinton's 80% less cited papers.

less impact, just less recognition. Good ideas may be hidden in less popular publications that need to be discovered and built upon before their true potential is revealed. This makes it obvious that newer publications cannot yet be effectively judged for high impact by citation count. Instead, one has to look for methodological different and possibly currently unpopular approaches. We will investigate this further with a qualitative analysis in the next section.

4 Successful ML research papers and their qualitative characteristics

Beyond the introduced metrics, we investigated what makes some of these most cited papers so significant. A good starting point for papers to look at are the publications of the co-author network from Figure 2 or the citation graph as aggregated by Ciriello [40]. Other sources are the 'test of time awards' awarded by some conferences such as NIPS.

We begin with the NIPS 2017 test of time award for Rahimi's paper on random features for kernel machines [41]. In his presentation [42] he makes some interesting remarks about the significance of his work. Besides their scientific contribution, they also carried out marketing during the conference by distributing flyers in order to create awareness. This was important because at the time access to machine learning frameworks was very limited. Not many of the current frameworks such as Scikit-Learn [43], Theano [44] or TensorFlow [45] were in existence. During the NIPS 2017 keynote [46] this claim was reinforced by the statement that not only is the availability of such frameworks driven by research, but also drives research itself.

Furthermore, we analyzed 12 publications [14, 16, 23, 31, 36, 47–53] of Jürgen Schmidhuber, Geoffrey Hinton and Yoshua Bengio. We picked the three most cited publications for each author in addition to publications with more than 5 thousand citations. The publications range from 1986 to 2015, at the time of writing less than three years ago, and include literature reviews, applications, as well as new methodologies. Unsurprisingly, as previously noted in the quantitative section, literature reviews are highly cited because it makes citing easier and allows the reader to follow up on the topic quicker. On the downside though, it may give attribution in the form of citations to the wrong authors.

Of the 12 publications we analyzed, 4 of them are literature reviews.

Most of the papers are not application papers but introduce new methodologies of learning in deep neural networks. The two exceptions are the deep ImageNet classifier [23] and the multi-column deep neural networks for image classification [51]. In both cases, breakthrough results have been achieved through the use of previous and more fundamental work that also has been cited highly but did not contain many experimental results. These papers achieved large impact by finally showing that previous research was not in vain and future research both built upon their success as well as compared new results with these landmark papers. One example is the series of ImageNet image classification competition papers [54–56] that came after Krizhevsky, Sutskever, and Hinton [23]. All other papers have a large contribution in the form of a new methodology for learning representations such as backpropagation [49] or regularization in the form of dropout [47]. One exclusion is Hinton's paper 'The appeal of parallel distributed processing' [50] that is focused on giving cognitive and psychological evidence why parallel processing such as done by neural networks is beneficial. In conclusion, we have direction-setting papers, fundamental new methodologies that only sometimes can show practical success straight away, and application papers that are relevant because of their contribution in showing previously theoretically introduced methodologies actually work well in larger scales.

Most of these papers also measure their success using metrics and display their superiority over previous approaches by showing that they work better on particular datasets. Most notably, their results were better by a large margin, e.g. on ImageNet [23], compared to many other less often cited papers that only show small improvements. To put this into perspective, the official ImageNet publication [57] which is usually cited whenever the dataset is used in a paper's experiments has been cited over 5 thousand times to this date. Simply improving upon other papers by achieving a better empirical result, therefore, is not a good indicator of the research's impact. Also, none of the reviewed papers invented their method entirely from scratch. To the contrary, most papers build on previous work and differentiate themselves from making small but very significant changes or making a previous idea finally work in practice. This is evidence for the previously stated hypothesis [26] that the field of deep learning is driven by engineering efforts. Thematically, all reviewed papers dealt with questions about how to effectively train neural networks both in the supervised and unsupervised context with the help of new algorithms such as backpropgation [49], regularization with dropout [47], autoencoders [48], deep belief networks [36], recurrent neural networks with large time-horizons [31] and practical considerations for efficient learning [23, 51]. The reviewed papers have been cited mostly because its methods are used in practice, new methods heavily use its introduced methodology, their contribution was a huge improvement in a certain area of application or the literature review was a useful target for citation.

Another aspect we intended to investigate is the question whether public media appearance of papers or hype coincides with high citation counts. It appears that the open advertisement of new research over platforms such as blogs and twitter is a phenomenon of the past few years and therefore not applicable to the reviewed publications.

4.1 Collaboratively curated lists

One source of deep learning papers with impact is a curated list on the platform GitHub [58]. With about 14 thousand GitHub stars at the time of writing it is one of the most popular sources for paper worth reading since 2012. Most notably, it focuses on collaborative decision-making and a fixed budget of 100 papers. Whenever a new paper is added, an old one must go to keep the number of papers constant. The authors also specify the criteria that papers must satisfy in order to be considered. New papers are accepted only by discussion and older papers require at least 200 citations per year. Generally, papers must be seminal and not limited to a field of application in machine learning research but applicable to a wide range of areas. Because GitHub is a collaborative platform we can simply inspect the reasoning for the inclusion of certain titles beyond citation count. A paper on deep residual networks [59], for instance, got accepted into the list due to showing empirical success that is applicable to a wide range of deep learning applications and appearing in lecture notes of Stanford University. Other indicators for inclusion are best paper awards from prestigious conferences [60].

5 Discussion

In this work, we mainly focused on qualitative analysis and citation count. We were able to show that highly cited work indeed had great impact both on academia, as well as industry and therefore society. Nevertheless, because the reverse - that less cited work had no impact - is not true, it is conceivable that better measures could be designed and used. Best paper awards and test of time awards may fill this gap but require large efforts of few individuals. A solution to this could be a further rise of collaboratively curated lists such as the one we investigated on GitHub, blogging, Twitter and other means of digital communication. Other options are ArXiv Sanity [61] that recommends publications based on other user's preferences. Still, these curated lists are highly dependent on the interest group that designs and maintains them. For instance, it has been found that ArXiv Sanity is much more focused on applications compared to conferences [62, 63]. Also, it is questionable whether these collaborative lists will take note of breakthrough ideas earlier than conventional research does.

Another question revolves around how the academic system has to be designed in order to promote impactful papers. As we have seen, many later very successful ideas were rather unpopular during their inception. This should encourage research groups, conferences, and journals to nurture crazy and currently unpopular ideas, possibly also through the introduction of a 'Crazy Work Award'.

We leave several frontiers to future investigation. Best paper awards could be qualitatively analyzed, differences between conferences could be highlighted, and the transition of research becoming basic knowledge taught at universities could lead to further insights. Additionally, the initial acceptance rate at conferences or journals of later to be discovered breakthrough papers might be of interest.

6 Conclusion

We have analyzed the co-author network of three important deep learning researchers and their publications. It was evident that just like in other areas of research, the most cited publications are distributed among a very small number of authors. We analyzed some of the particularly successful papers and observed the majority of citations only decades later after large-scale practical success of these methods was evident. Great ideas in deep learning research therefore often require perseverance for long periods of time and success often depends on whether the approach can be scaled up to large problems that are useful for consecutive research or industry adoption. The case of backpropagation and LSTMs also showed that not the mainstream research established itself but the novel ideas that were not generally accepted. When analyzing 12 highly cited papers it also became evident that working on applications only leads to high impact in the form of citations and adoption if it uses novel techniques that improve over existing work by a large margin. Small improvements are quickly surpassed. None of the application papers with large numbers of citations solely applied existing techniques to a new application. Therefore, to maximize impact within the research community, effort should be focused on general learning algorithms over applications. Nevertheless, it is crucial to demonstrate the effectiveness of these learning algorithms on large-scale problems to be recognized. Through experiments we showed that many contributions even from famous authors are barely cited, rendering deep learning research a form of trial and error. If these findings were to be applied to a researcher's style of research it is noteworthy that many of our observations might not generalize well into the future. Nevertheless, the preceding analysis could be a good starting point for the direction of future research endeavours in the field of machine learning.

References

[1] Michael Chui. "Artificial intelligence the next digital frontier?" In: *McKinsey and Company Global Institute* (2017), p. 47.

[2] Robbie Allen. *NIPS Accepted Papers Stats*. 2017. URL: `https://unsupervisedmethods.com/nips-accepted-papers-stats-26f124843aa0` (visited on 03/17/2018).

[3] *arXiv.org e-Print archive*. 2013. URL: `https://arxiv.org/` (visited on 03/09/2018).

[4] Andrej Karpathy. *A Peek at Trends in Machine Learning*. 2017. URL: `https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106` (visited on 03/09/2018).

[5] Lutz Bornmann. "Measuring impact in research evaluations: a thorough discussion of methods for, effects of and problems with impact measurements". In: *Higher Education* 73.5 (2017), pp. 775–787. ISSN: 0018-1560. DOI: `10.1007/s10734-016-9995-x`.

[6] Lutz Bornmann. "Scientific Peer Review". In: *Annual Rev. Info. Sci & Technol.* 45.1 (Jan. 2011), pp. 197–245. ISSN: 0066-4200. DOI: `10.1002/aris.2011.1440450112`. URL: `http://dx.doi.org/10.1002/aris.2011.1440450112`.

[7] Kiri Wagstaff. "Machine Learning that Matters". In: (2012). arXiv: 1206.4656.

[8] Diana Hicks et al. "Bibliometrics: The Leiden Manifesto for research metrics". In: *Nature* 520.7548 (Apr. 2015), pp. 429–431. ISSN: 0028-0836. DOI: 10.1038/520429a. URL: http://www.ncbi.nlm.nih.gov/pubmed/25903611.

[9] Hadas Shema, Judit Bar-Ilan, and Mike Thelwall. "Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics". In: *Journal of the Association for Information Science and Technology* 65.5 (May 2014), pp. 1018–1027. ISSN: 23301643. DOI: 10.1002/asi.23037. arXiv: 0803.1716. URL: http://doi.wiley.com/10.1002/asi.23037.

[10] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[11] Simon Haykin and Neural Network. "A comprehensive foundation". In: *Neural networks* 2.2004 (2004), p. 41.

[12] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.

[13] Christopher M Bishop. "Pattern recognition and machine learning". In: *Springer-Verlag New York* 1 (2006).

[14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[16] Juergen Schmidhuber. "Deep Learning in Neural Networks: An Overview". In: (Apr. 2014). DOI: 10.1016/j.neunet.2014.09.003. arXiv: 1404.7828. URL: http://arxiv.org/abs/1404.7828.

[17] Jürgen Schmidhuber. *Who Invented Backpropagation?* 2015. URL: http://people.idsia.ch/~juergen/who-invented-backpropagation.html (visited on 03/16/2018).

[18] H J Kelley. "Gradient Theory of Optimal Flight Paths". In: *ARS Journal* 30.10 (1960), pp. 947–954.

[19] Arthur E Bryson Jr. and W F Denham. *A steepest-ascent method for solving optimum programming problems*. Tech. rep. BR-1303. Raytheon Company, Missle and Space Division, 1961.

[20] A E Bryson. "A gradient method for optimizing multi-stage allocation processes". In: *Proc. Harvard Univ. Symposium on digital computers and their applications*. 1961.

[21] Werner Marx and Lutz Bornmann. "How accurately does Thomas Kuhn's model of paradigm change describe the transition from the static view of the universe to the big bang theory in cosmology?" In: *Scientometrics* 84.2 (2010), pp. 441–464.

[22] Lutz Bornmann, Rüdiger Mutz, and Hans-Dieter Daniel. "Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/2012". In: *Journal of the Association for Information Science and Technology* 64.8 (2013), pp. 1649–1658.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances In Neural Information Processing Systems* (2012), pp. 1–9. ISSN: 10495258. DOI: http://dx.doi.org/10.1016/j.protcy.2014.09.007. arXiv:

1102.0183. URL: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[24] OrganicIrradiation. *Python library scholarly.* 2013. URL: https://github.com/OrganicIrradiation/scholarly.

[25] Bruno Latour. *Science in action: How to follow scientists and engineers through society.* Harvard university press, 1987.

[26] Andrew Beam. *Deep Learning 101 - Part 1: History and Background.* 2017. URL: https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html (visited on 04/03/2018).

[27] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), p. 533.

[28] James Somers. *Is AI Riding a One-Trick Pony? - MIT Technology Review.* 2017. URL: https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/ (visited on 03/18/2018).

[29] Qing Ke et al. "Defining and identifying Sleeping Beauties in science". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.24 (June 2015), pp. 7426–31. ISSN: 0027-8424. DOI: 10.1073/pnas.1424329112. arXiv: 1505.06454. URL: http://www.ncbi.nlm.nih.gov/pubmed/26015563.

[30] Anthony F J Van Raan. "Sleeping beauties in science". In: *Scientometrics* 59.3 (2004), pp. 467–472.

[31] S Hochreiter and J Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[32] A Graves and J Schmidhuber. "Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks". In: *Advances in Neural Information Processing Systems (NIPS) 21.* Cambridge, MA: MIT Press, 2009, pp. 545–552.

[33] M Liwicki et al. "A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks". In: *Proceedings of the 9th International Conference on Document Analysis and Recognition.* Sept. 2007.

[34] Alex Graves, Abdel-Rahman Mohamed, and Geoffrey E Hinton. "Speech recognition with deep recurrent neural networks". In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE. 2013, pp. 6645–6649.

[35] Per O. Seglen. "The skewness of science". In: *Journal of the American Society for Information Science* 43.9 (Oct. 1992), pp. 628–638. ISSN: 10974571. DOI: 10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0. arXiv: 0912.4188.

[36] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. "A Fast Learning Algorithm for Deep Belief Nets". In: *Neural Computation* 18.7 (May 2006), pp. 1527–1554. ISSN: 0899-7667.

[37] Lutz Bornmann, Félix de Moya Anegón, and Loet Leydesdorff. "Do Scientific Advancements Lean on the Shoulders of Giants? A Bibliometric Investigation of the Ortega Hypothesis". In: *PLoS ONE* 5.10 (Oct. 2010). Ed. by Pedro Antonio Valdes-Sosa, e13327. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0013327. URL: http://dx.plos.org/10.1371/journal.pone.0013327.

[38] John P. A. Ioannidis, Kevin W. Boyack, and Richard Klavans. "Estimates of the Continuously Publishing Core in the Scientific Workforce". In: *PLoS ONE* 9.7 (July 2014). Ed. by Luís A. Nunes. Amaral, e101698. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0101698. URL: http://dx.plos.org/10.1371/journal.pone.0101698.

[39] Karl Popper. *The Logic of Scientific Discovery.* 1959, p. 513. ISBN: 0203994620. URL: http://strangebeautiful.com/other-texts/popper-logic-scientific-discovery.pdf.

[40] Daniele Ciriello. *Five Hundred Deep Learning Papers, Graphviz and Python.* URL: https://dnlcrl.github.io/projects/2015/10/10/500-deep-learning-papers-graphviz-python.html (visited on 03/17/2018).

[41] Ali Rahimi and Ben Recht. "Random features for large-scale kernel machines". In: *Advances in neural information ...* 1 (2007), pp. 1–8. ISSN: 0033-6599. DOI: 10.1.1.145.8736. URL: https://people.eecs.berkeley.edu/~brecht/papers/07.rah.rec.nips.pdf.

[42] Ali Rahimi. *NIPS 2017 Test-of-time award presentation - YouTube.* 2017. URL: https://www.youtube.com/watch?v=Qi1Yry33TQE (visited on 03/16/2018).

[43] F Pedregosa et al. "Scikit-learn: Machine Learning in {P}ython". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[44] Rami Al-Rfou et al. "Theano: A {Python} framework for fast computation of mathematical expressions". In: *arXiv e-prints* abs/1605.0 (May 2016). URL: http://arxiv.org/abs/1605.02688.

[45] Martín Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: (). URL: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45166.pdf.

[46] Yee Whye Teh. *On Bayesian Deep Learning and Deep Bayesian Learning (NIPS 2017 Keynote) - YouTube.* 2017. URL: https://www.youtube.com/watch?v=9saauSBgmcQ%7B%5C&%7Dt=2144s (visited on 03/20/2018).

[47] Nitish Srivastava et al. "Dropout: A simple way to prevent neural networks from overfitting". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1929–1958.

[48] Geoffrey Hinton and Ruslan Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (2006), pp. 504–507.

[49] D E Rumelhart, G E Hinton, and R J Williams. "Learning Internal Representations by Error Propagation". In: *Parallel Distributed Processing.* Ed. by D E Rumelhart and J L McClelland. Vol. 1. MIT Press, 1986, pp. 318–362.

[50] James L McClelland, David E Rumelhart, and Geoffrey E Hinton. "The appeal of parallel distributed processing". In: *MIT Press, Cambridge MA* (1986), pp. 3–44.

[51] D C Ciresan, U Meier, and J Schmidhuber. "Multi-Column Deep Neural Networks for Image Classification". In: *IEEE Conference on Computer Vision and Pattern Recognition CVPR 2012.* 2012.

[52] Y LeCun et al. "Gradient-Based Learning Applied to Document Recognition". In: *Proceedings of the IEEE* 86.11 (Nov. 1998), pp. 2278–2324.

[53] Y Bengio. *Learning Deep Architectures for AI. Foundations and Trends in Machine Learning, V2(1).* Now Publishers, 2009.

[54] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[55] Christian Szegedy et al. "Scalable, high-quality object detection". In: *arXiv preprint arXiv:1412.1441* (2014).

[56] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.0 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[57] Olga Russakovsky et al. "Imagenet large scale visual recognition challenge". In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252.

[58] Terry Taewoong Um. *Awesome Deep Learning Papers.* 2018. URL: https://github.com/terryum/awesome-deep-learning-papers (visited on 03/17/2018).

[59] Kaiming He et al. "Identity Mappings in Deep Residual Networks". In: (Mar. 2016). ISSN: 16113349. DOI: 10.1007/978-3-319-46493-0_38. arXiv: 1603.05027. URL: http://arxiv.org/abs/1603.05027%20https://arxiv.org/pdf/1603.05027.pdf.

[60] Moritz Hardt and Oriol Vinyals. "Understanding Deep Learning Requires Re- Thinking Generalization". In: *International Conference on Learning Representations 2017* (Nov. 2017), pp. 1–15. ISSN: 10414347. DOI: 10.1109/TKDE.2015.2507132. arXiv: arXiv:1611.03530v1. URL: http://arxiv.org/abs/1611.03530.

[61] Andrej Karpathy. *ArXiv Sanity.* 2015. URL: http://www.arxiv-sanity.com/ (visited on 05/03/2018).

[62] Carlos Perez. *Ten Deserving Deep Learning Papers that were Rejected at ICLR 2017.* 2017. URL: https://medium.com/intuitionmachine/eight-deserving-deep-learning-papers-that-were-rejected-at-iclr-2017-119e19a4c30b (visited on 05/03/2018).

[63] Andrej Karpathy. *ICLR 2017 vs arxiv-sanity.* 2017. URL: https://medium.com/@karpathy/iclr-2017-vs-arxiv-sanity-d1488ac5c131 (visited on 05/03/2018).