ISSN 1060-992X, Optical Memory and Neural Networks, 2018, Vol. 27, No. 2, pp. 65–80. © Allerton Press, Inc., 2018.

Review of State-of-the-Art in Deep Learning Artificial Intelligence

V. V. Shakirov[a, b], K. P. Solovyeva[a, b], and W. L. Dunin-Barkowski[a, b, *]

[a] *Scientific Research Institute of System Analysis, Moscow, Russia*

[b] *Moscow Institute of Physics and Technology, Russia*

**e-mail: wldbar@gmail.com*

Received December 5, 2017; in final form, March 1, 2018

Abstract—The current state-of-the-art in Deep Learning (DL) based artificial intelligence (AI) is reviewed. A special emphasis is made to compare the level of a concrete AI system with human abilities to show what remains to be done to achieve human level AI. Several estimates are proposed for comparison of the current "intellectual level" of AI systems with the human level. Among them is relation of Shannon's estimate for lower bound on human word perplexity to recent progress in natural language AI modeling. Relations between the operation of DL constructions and principles of live neural information processing are discussed. The problem of AI risks and benefits is also reviewed based on arguments from both sides.

Keywords: deep learning, artificial intelligence, natural language processing, residual networks, generative adversarial networks, Pavlov Principle

DOI: 10.3103/S1060992X18020066

1. INTRODUCTION

Recent progress in deep learning (DL) algorithms for artificial intelligence (AI) has improved the AI field so much that it is worth to review the overall picture that we have now. One of the important questions, which arise in this field, is an estimation of the time when the human-level AI will be created [1]. Certainly it's very hard to predict future of science. In this article the recent progress in DL is reviewed. This progress has some regularities which enable its extrapolation into the future. Such extrapolation is in favor of reasonably high probability of human level AI in next 5 to 10 years or sooner.

It has been argued that emergence of human-level AI isn't automatically good for humanity. On one hand, there are arguments that humans will get beneficial partnership with superhuman-clever AIs, or that those AIs would really care about humans, like allowing us appropriate access to mineral resources and agriculture fields of the planet [2]. On the other hand, these arguments might be futile due to their vagueness and the supposed very long time distance from now [3]. However, times are changing. Superhuman-level AI might be quite near as well [1].

The field is hot and most of the works, cited in this review, are from arXiv, not from journals. The paper most similar to the current article has been written in 2013 by Katja Grace [4]. Another review with more details on deep history of deep learning has been written in 2014 by Jurgen Schmidhuber [5].

The structure of the article is as follows. After Introduction, in Sections 2 and 3, we review state-of-the-arts in natural language processing and sensory image processing. In Sections 4 and 5 progress in DL methods is discussed. Sections 6–8 address the issues in general AI, which seem to be important, based on cognitive sciences and neurosciences. Also, in Section 8 we discuss the feedback influence of DL success onto ideas and methods of neuroscience. Here it is also argued that for both fields—AI and Neurosciences, the important role might be attributed to Pavlov Principle [6]. This is a generalization of original I.P. Pavlov early intuitive guess about the importance for physiology and psychology of the conditioned reflexes (discovered in his laboratories) [7] and the current success story of DL. In Sections 9, some new and prospective approaches are reviewed. Many of them might revolutionize modern deep learning but we don't know yet which ones. In Section 10, the predictions of professionals about when human level AI would be developed are discussed. In Section 11, a brief review of AI safety research is given. Conclusions are presented in Section 12.

Table 1. Best perplexity scores

Dataset	Perplexity	Link
Wikipedia english corpus snapshot 2014/09/17 (1.5B words)	27.1	[66]
1B word benchmark (shuffled sentences)	24.2	[105]
OpenSubtitles (923M words)	17	[83]
IT Helpdesk Troubleshooting (30M words)	8	[83]
Movie Triplets (1M words)	27	[122]
PTB (1M words)	62.34	[160]

Table 2. Best perplexity scores, single model

Number of hidden neurons	Perplexity in [6]	Perplexity in [105]
256	38	–
512	–	54
1024	27	–
2048	–	44
4096	–	–
8192	–	30 24.2 (ensemble)

2. NATURAL LANGUAGE PROCESSING

We begin with review of state-of-the-arts in natural language modeling. For estimation of (probabilistic) modeling quality the perplexity measure or just *perplexity* is currently in use. Numerical value of perplexity is the mean value of $1/P(t, t+1)$, where $P(t, t+1)$ is the probability that the model chooses the correct word $w(t+1)$ in the moment $t+1$, given all the words $w(t')$, which were present on or before the moment t. Table 1 shows perplexity values for the best current AI programs.

Why perplexity matters?

If neural net chooses inconsistent next word (wrong in any sense: logical, syntactic, pragmatic) then it means that the system has not been correctly tuned to the text, to which it reacts with generation of the next word. When neural chat bots are trained to produce words with low perplexity in relation to the training corpus of texts, they become capable to write coherent stories, to answer intelligibly, with common sense, to questions, related to the recently loaded information, to reason in a consistent and logical way, etc. [8].

Shannon estimated lower and upper bounds of human perplexity to be 0.6 and 1.3 bits per character [9]. Strictly speaking, applying formula (17) in his work gives lower bound equal 0.648 which he rounded. Given average word length of 4.5 symbols and including spaces (as Shannon included them in his study) gives us an estimate for lower bound on human-level word perplexity as $11.8 = 2^{0.648 \times 5.5}$. This lower bound might be much less than real human perplexity [10].

What are reasonable predictions for perplexity improvements in the nearest future? Table 2 shows results of works [11] and [12] and their dependence on the number of hidden neurons, hs. From the data, it's quite reasonable to extrapolate that algorithm [11] for hs = 4096 might give pplx < 20, and hs = 8192 might give pplx < 15 while ensemble of models with hs = 8192 trained on 10B words might give perplexity well below 10. Nobody can tell now what kind of common sense reasoning would such a neural net have.

Another source of improvement may come from solving some discrepancy in what kind of perplexity is optimized [13], [14]. Here also the ways are proposed to partially solve that problem using adversarial learning. The work [15] demonstrates impressive advantages of adversarial paradigm. Wasserstein GAN is another good direction of research [16].

Human BLEU score for Chinese to English translation on MT03 dataset is 35.76 [17]. In recent article [18] neural network gets 40.06 BLEU on the same task and dataset. They took state-of-the-art [19] "GroundHog" network and replaced maximum likelihood estimation with their own MRT criterion, which increased BLEU from 33.2 to 40.06. Here is a quote from abstract: "Unlike conventional maximum likelihood estimation, minimum risk training is capable of optimizing model parameters directly with

Table 3. Performance improvement on several tasks

Year	ImageNet top-5 error	ImageNet localization	PASCAL VOC detection [161]
2011	25.77%	42.5%	
2012	15.31%	33.5%	
2013	11.20%	29.9%	<50%
2014	6.66%	25.3%	63.8%
2015	3.57%	9.0%	76.4%
2016	2.99%	7.7%	88.4%

respect to evaluation metrics". Another impressive improvement comes from improving translation with monolingual data [20]. Modern neural nets translate text ~1000 times faster than humans do [21]. The most surprising recent result in language translation is translation without using the parallel texts [22]. The author's approach yields BLEU score for the language pairs (French-English) and (German-English) without using a single word of parallel texts. Indeed, the result might be due to the fact that, to a certain degree, English is a kind of a hybrid between French and German. Nevertheless, the result is absolutely amazing.

3. SENSORY SIGNALS PROCESSING

Table 3 clearly illustrates fast rate of progress in computer vision.

"Identity mappings in deep residual networks" [23] reach 5.3% top-5 error in single one-crop model while human level is reported to be 5.1% [24]. In "deep residual networks" [25] one-crop single model gives 6.7% but ensemble of those models with Inception gives 3.08% [26]. Another big improvement comes from "deep networks with stochastic depth" [27]. There was reported ~0.3% error in human annotations to ImageNet [24], so real error on ImageNet would soon become even below 2%. Human level is overcome not only in ImageNet classification task (see also an efficient implementation of 21841 classes ImageNet classification [28]) but also on boundary detection [29]. Video classification task on SPORTS-1M dataset (487 classes, 1M videos) performance improved from 63.9% [30] (2014) to 73.1% [31] (March 2015). See also [32].

CNNs outperform humans also in terms of speed being ~1000x faster than human [33] (note that times there are given for batches) or even ~10000× times faster after compression [34]. Video processing 24fps on AlexNet demands just 82 Gflops and GoogleNet demands 265 Gflops. Here's why. The best benchmark in [33] gives 25ms feedforward and 71ms total time for 128 pictures batch on NVIDIA Titan X (6144 Gflops), so for 24fps video real-time feedforward processing we need 6144 Gflops * 24/128 * 0.025 = 30 Gflops. If we want to learn something using backprop than we need 6144 Gflops * 24/128 * 0.071 = 82 Gflops. Same calculations for GoogleNet give 83 Gflops and 265 Gflops respectively.

Neural nets can answer questions based on images [35]. Using similar method as in [36] the equivalent human age of net [35] can be estimated as 6.2 years old (submitted 4 Mar 2016), while [37] was 5.45 years (submitted 7 Nov 2015), [36] was 4.45 years (submitted 3 May 2015). See Appendix for details of age estimates. Also, nets can describe images with sentences, in some metrics even better than humans can [38]. Beside video=>text [39–41], there are some experiments to implement drawing pictures, based on text [42–44].

Speech recognition closely resembles and follows computer vision. Table 4 shows that here, AI hasn't surpassed human level yet but it's clearly seen that we have all chances to see it in 2017. The rate of improvement is very fast. For example, Google reports it is word error rate dropped from 23% in 2013 to 8% in 2015 [45].

Multimodal learning is used in [46–49] to improve video classification related tasks. Unsupervised multimodal learning is used in grounding of textual phrases in images [50] using attention-based mechanism so that different modalities supervise one another. In "neural self-talk" [51] a neural network sees a picture, generates questions based on it and answers those questions itself.

Table 4. Speech recognition performance errors

Year	CHiME noisy	VoxForge European	WSJ eval'93	LibriSpeech test-other	Citation
2014	67.94%	31.2%	6.94%	21.74%	[162]
2015	21.79%	17.55%	4.98%	13.25%	[163]
human	11.84%	12.76%	8.08%	12.69%	[163]

4. REINFORCEMENT LEARNING

The very sound Al results with reinforcement learning have been obtained in playing an ancient and highly respected for its intellectual values the game of Go [52]. The latter operates in its seemingly very simple but in fact extremely complicated world, which is just a board of 19 × 19 lines with stones and a bunch of simple rules. The final AI result with Go is absolutely impressive. Computer learns to play the game by self-learning from the beginning, and defeats all human players and all the programs, which took into account huge amounts of human experience.

Go and Atari playing agents choose one action from a finite set of possible actions. There are also articles on *continuous* reinforcement learning where action is a vector [53–61].

If we enable the successful Go playing programs with continuous reinforcement learning (and if we manage to make it work well on hard real-world tasks) than it would be real AGI in a real world. Also, we can begin with pre-training it on virtual videogames [62]. Videogames often contain even much more interesting challenges than real life gives for an average human. Millions of available books and videos contain the concentrated life experience of millions of people. While Go programs present successful examples of pairing modern RL with modern CNN, RL can be combined with neural chat bots and reasoners [63, 64].

The success of reinforcement learning algorithms in technological applications enables one to suggest a plausible solution to one of the old evolutional enigmas: what is the function of dreams? The hypothesis is based on the fact that before humans started to talk there were no arguments to them to make distinction between life events and events in dreams. That means that human behavior sometimes could be controlled by the latter. On the other hand dreams are more or less realistic, but practically never follow logics of everyday life. In reinforcement learning theory [65] one of the principle ideas is using of the "epsilon-greedy" strategy. That means that you almost all the time take "greedy" actions (which seem to provide maximal expected reward) and with a (small) probability epsilon make a "random" action. Decision

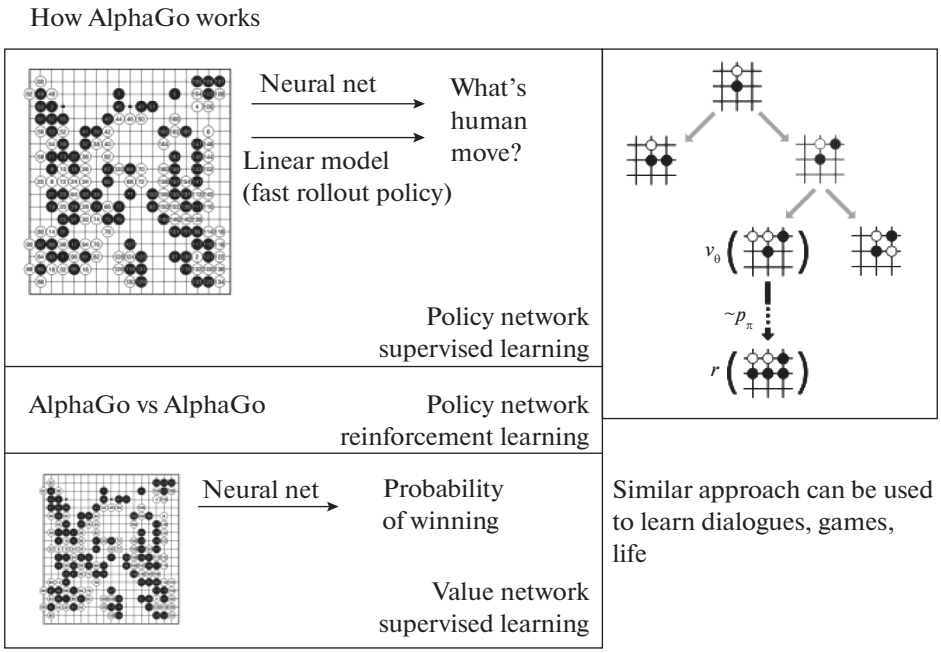

Fig. 1. How AlphaGo works [137].

based on events in dream can play this random role. It seems that thus formulated dream function hypothesis has not been proposed earlier [66].

5. UNSUPERVISED LEARNING

DCGAN [67] generates reasonable pictures [68]. Language generation models which minimize perplexity are unsupervised and were improved very much recently (Table 1). "Skip-thought vectors" [69] generate vector representation for sentences allowing to train linear classifiers over those vectors and their cosine distances to solve many supervised problems at state-of-the-art level. Generative nets gained deep new horizons at the edge of 2013–2014 [70, 71]. Recent work continues progress in "computer vision as inverse graphics" approach [72].

Another recent breakthrough is Wasserstein GAN [16] which improved stability of GAN training making it much easier to train GAN with wider range of hyperparameters.

6. IS EMBODIMENT CRITICAL?

People with tetra-amelia syndrome [73] have neither hands nor legs from their birth and they still manage to get a good intelligence. For example, Hirotada Ototake [74] is a Japanese sports writer famous for his bestseller memoirs. He also worked as a school teacher. Nick Vujicic [75] has written many books, graduated from Griffith University with a Bachelor of Commerce degree, he often reads motivational lectures. Prince Randian [76] spoke Hindi, English, French, and German. Modern robots have better embodiment.

There are quick and good drones, there are quite impressive results in robotic grasping [77]. It's argued that embodiment might be done in virtual world of videogames [62]. Among 8 to 18-year-olds average amount of time spent with TV/computer/video games etc., is 7.5 hours a day [78]. According to another study [79] UK adults spend an average an 8 hours and 40 minutes a day on media devices. Humans develop commonsense intelligence very fast. When you are 2 years old you barely can do something that current AI can't. When you're 4 years old you have common sense, you can learn from texts and conversations. However, 2 years are just 100 weeks. If we exclude sleep we get ~50 weeks. The 24 fps video input of 50 weeks might be processed in 10 hours on pretrained AlexNet using four Titan X.

7. ARGUMENTS FROM NEUROSCIENCE

Roughly speaking [80], [81] 15% of human brain is devoted to low-level vision tasks (occipital lobe). Another 15% are devoted to image and action recognition (somewhat more than a half of temporal lobe). Next 15% are devoted to objects detection and tracking (parietal lobe). Then, next 15% are devoted to speech recognition and pronunciation (Brodman Areas 41, 42, 22, 39, 44, parts of 6, 4, 21) and, also, 10% are devoted to reinforcement learning (orbitofrontal cortex and part of medial prefrontal cortex). Taken together, the listed cortex parts comprise about 70% of human brain.

Judging on these facts, we can say that modern neural networks work at about human level for these 70% of human brain.

For example, CNNs make 1.5× less mistakes than humans at ImageNet while acting about 1000× times faster than a human.

One can claim that unexplained in function remain just 30% of human brain cortex.

According to modern micro-anatomical and neurophysiological studies, human cortex has the similar structure throughout its entire surface [82]. It's just 3mm thick mesh of neurons functioning on the same principles throughout all the cortex. There is likely no big difference between how prefrontal cortex works and how other parts of cortex work. There is likely no big difference in their speed of calculations, in complexity of their algorithms. It would be somewhat strange if modern deep neural networks can't solve remaining 30% in several years.

The standard neuroanatomical view of human brain subdivision into 47 Brodman Areas (BA) is given in Fig. 3. About 10% of the entire cerebral cortex area belong to low-level motorics (BAs 6,8). Robotic hands are not very dexterous. However people having no fingers from their birth have problems with fine motorics but still develop normal intelligence [73], see also the section "Is embodiment critical?". Also, one of functions of a part of brain cortex, called DLPFC, is attention which is actively used now in LSTMs.

The only part which still lacks near human-level performance are BAs 9, 10, 46, 45 which together occupy only 20% of human brain cortex. These areas are responsible for complex reasoning, complex tools

usage, complex language. However, "A neural conversational model" [83], "Contextual LSTM..." [84], "playing Atari with deep reinforcement learning" [85], "mastering the game of Go..." [52], and numerous other already mentioned articles have recently begun to really attack this problem.

There seem to be no particular reasons to expect these 30% to be much harder than other 70%. Less than 3 years has gone from AlexNet winning ImageNet competition to surpassing human level. It's reasonable to expect the same <3 years gap from "A neural conversational model" winning over Cleverbot to human-level reasoning. After all, there are much more deep learning researchers now with much more knowledge and experience, there are much more companies interested in DL.

Extensive survey of approaches and perspectives to compose Human Level AI from the point of view of Computational Cognitive Sciences is given in [86]. Also, this publication is accompanied by critics and discussion of many authors, as this is practiced by Behavioral and Brain Sciences.

8. PROGRESS IN BRAIN REVERSE ENGINEERING. PAVLOV PRINCIPLE

Detailed connectome deciphering has begun to really succeed recently with creation of multi-beam scanning electron microscopy [87], which enable labs to get a grant for deciphering the detailed connectome of $1 \times 1 \times 1$ mm^3 of rat cortex [88] after the proof-of-principle deciphering of $40 \times 40 \times 50$ mcm of brain cortex with $3 \times 3 \times 30$ nm resolution was achieved [89].

Of particular importance for understanding connections between success of DL systems and real neural systems is the work of Lillicrap, T. P. et al. [90]. These authors have demonstrated that error backpropagation (BP) is not necessary for the tuning of DL systems. This is crucial because the BP cannot be implemented in live neural systems. Thus, there came understanding that "the secrets" of real brains cognitive abilities might be due to the same mechanisms, which provide smartness of DL neural networks. Therefore, for construction of "artificial brain" of any intellectual power no principal difficulties are expected now. It should be emphasized at this point that currently arXiv delivers much more timely information than even the best journals. The journal version of [90] appeared in very prestigious Nature Communications two years later than the original publication [91].

It should be also noted that all computational experimental data along with elements of their mathematical foundation (the true mathematical theory for efficiency of DL systems is yet to come) in fact demonstrate that Ivan P. Pavlov had got a correct guess back in 1923 that myriads of elementary conditioned reflexes can be responsible for the human intellectual abilities [7]. Pavlov's initial foresight had been detailed at the level of conditioned reflexes synaptic mechanisms by Donald Hebb [92]. Recently (April, 2016), the general formulation had been proposed that unifies the classical ideas/prescience of Pavlov and Hebb and the current progress in deep learning [6]. It is made in form of a "principle" or word formula, naturally named Pavlov Principle (PP) to mark the Pavlov's envision as following.

Pavlov Principle. A neural network, in which the strength of each inter-neuronal connection, gradually changes as a function of locally distributed vector error signals and the activity states of connected neurons, gradually evolves to error-free functioning [6].

This principle just generalizes on the current experimental knowledge and yields clues for understanding live brains and for construction of artificial neural intelligence systems. It is important that in December, 2016, the new version of DL has been published. It was named "Direct Feedback Alignment" [93]. This method exactly corresponds to PP, as here the error signal goes straight from the site(s) of error detection to all synapses in the network, without propagation of error signals from layer to layer in BP (and as still takes place in "Random Feedback alignment" [90]). Also important for the work of PP is another important feature of large neural networks (containing 300+ neurons) is randomness of interneuronal connections. This feature is also a must for initial conditions of interlayer matrices in all DL networks. The needed randomness is not a self-sustained goal, but it is just the best way to implement in neural networks non-trivial, complex functions, which are needed for information transformations [94, 95]. Mathematical proof of PP is hardly possible in near future as it should be valid for any "reasonable" neuron or neuron model, the latter notion falls out the mathematical scope. Nevertheless, attempts to imply physical and mathematical reasoning to explain why DL works are continuing [96–99].

In [100] the standard BP is characterized as a construction with weight symmetry, as in BP the signals, connected with error detection are propagated backward from output, using the same connections (with the same weights) as direct signals, which implement processing of the input signals. All other means of bringing the error information from output inside the neural network is characterized in [100] as non-symmetric BP. However, in PP, as well as in [93], errors are not propagated from layer to layer but are directly distributed among synapses in all layers of the network. It should be noted that BP provides a con-

venient tool for tuning artificial neural networks and is widely used in ML (despite the fact that it is not "Biologically plausible").

Recently, STDP objective function has been proposed [101]. It's like an unsupervised objective function somewhat similar to what is used for example in word2vec. In [98] surveyed a space of polynomial local learning rules (learning rules in brain are supposed to be local) and found that backpropagation outperforms them. There are also online learning approaches which require no backpropagation through time, for example, [102]. Although they can't compete with conventional deep learning in ML, brain perhaps can use something like that given fantastic number of it's neurons and synaptic connections. The STDP is an experimentally characterized version of the idea of "Hebb synapse". The latter, in turn, presents in fact single neuron implementation of conditioned reflexes. So, results of [98, 101, 102], along with the BP itself, might be considered as established examples of PP realization.

9. PROSPECTIVE DIRECTIONS IN NEUROMORHIC AI TECHNOLOGIES

Recently, a flow of articles about memory networks and neural Turing machines made it possible to use arbitrarily large memories while preserving reasonable number of model parameters. Hierarchical attentive memory [103] (Feb2016) allowed memory access in $O(\log n)$ complexity instead of usual $O(n)$ operations, where n is the size of the memory. Reinforcement learning neural Turing machines [104] (May2015) allowed memory access in $O(1)$. It's a significant step towards realizing systems like IBM Watson on completely end-to-end differentiable neural networks and in order to improve Allen AI challenge results from 60% to the numbers close to 100%. One also can use bottlenecks in recurrent layers to get large memories using reasonable amount of parameters like in [12]).

Neural programmer [105] is a neural network augmented with a set of arithmetic and logic operations. It might be first steps toward end-to-end differentiable Wolphram Alpha realized on a neural network. The "learn how to learn" approach [106–108] has a great potential.

Recently, cheap stochastic variance reduced gradient (SVRG) method [109] was proposed. This line of work aims to use theoretically very much better converging gradient descent methods.

There are works which when succeed would allow training with immensely large hidden layers, see in "unitary evolution RNN" [110], "tensorizing neural networks" [111], "virtualizing DNNs..." [112].

"Net2net" [113] and "Network morphism" [114] allow automatically initialize new architecture NN using weights of old architecture NN to obtain instantly the performance of latter. Collections of pretrained models are accessible online [115–117] etc. It is, in fact, an initial stage of module based approach to neural networks. One just downloads pretrained modules for vision, speech recognition, speech generation, reasoning, robotics etc. and fine-tunes them on final task.

It's very reasonable to include new words in a sentence vector in a deep way. However modern LSTMs update cell vector when given new word in almost shallow way. This can be solved with the help of deep transition RNNs proposed in [118] and further elaborated in [119]. Recent successes in applying batch normalization to recurrent layers [120] and applying dropout to recurrent layers [121] might allow to train deep-transition LSTMs even more effectively. It also would help hierarchical recurrent networks like [40, 41, 122–124]. Recently, several state-of-the-arts were beaten with an algorithm that allows recurrent neural networks to learn how many computational steps should be taken between receiving an input and emitting an output [125]. Ideas from residual nets might also improve performance. For example, stochastic depth neural networks [27] enable increase of the depth of residual networks beyond 1200 layers while getting state-of-the-art results.

Memristors might accelerate neural networks training by several orders of magnitude and make it possible to use trillions of parameters [126, 127]. Quantum computing promises even more [128, 129]. Recently, the 51 (fifty one!) qubit quantum computer was demonstrated by Mikhail D. Lukin and his team [130].

Deep learning is easy. Deep learning is cheap. Best articles usually use no more than several dozens of GPU. For half billion dollars one can buy 64000 NVIDIA M6000 GPUs with 24 Gb RAM, ~7 teraflops each, including processors etc. to make them work. For another half billion dollars one can prepare 2000 highly professional researchers from those one million [131] enrollments on Andrew Ng's machine learning course on Coursera. So for a very feasible R&D budget for every big country or corporation, one gets two thousand professional AI researchers equipped with 32 best GPUs each. It's kind of investment very reasonable to expect during next years of explosive AI technologies improvement.

The difference between year 2011 and year ~~2017~~ is enormous. The difference between ~~2017~~ and 2023 would be even much more impressive because we have now 1-2 orders of magnitude more researchers and

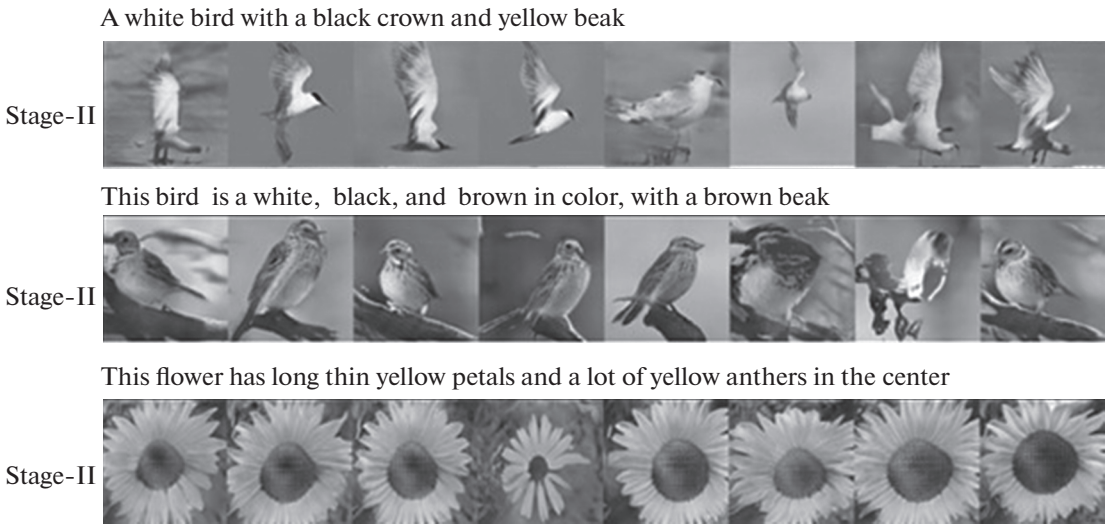

Fig. 2. Examples generated by GAN [164].

companies deeply interested in DL. The topics of neural networks based AI and machine learning applications now often appear in mass-media and at all levels of business applications [132]. The current market of the latter costs billions of dollars and involves numerous extremely diverse fields, from NLP (machine synchronous translation), banking customers telling, driver-free cars, drug design [133], etc.

Finally, it should be mentioned that some neuromorphic technologies might be based not on DL, but on special types of three layer perceptrons. In particular, currently, there are no doubts that the core structure in the cerebellum is a version of the three layer perceptron [134; and others]. The main distinctive feature of the cerebellar machinery is a huge number of "second layer perceptron cells", which act on one output. These cells are granule cells of the cerebellum. In human cerebellum their number constitutes about 70% of the total count of brain nerve cells. One functional module of the cerebellum includes one climbing fiber cell (located in inferior olives), ten (average number) Purkinje cells and up to 2000000 granule cells. This "equipment" provides flying abilities for those animals, who can fly (birds and bats). Other motor control functions of cerebellum are less evident. Since 1990-ies definite cognitive functions were ascribed to cerebellum. Although their existence was implicated in many studies, their explicit characterization have been lacking. On the ground of pure analogy with the cerebellar role in flying, cited above, it has been suggested that the role of cerebellum in human cognition is to provide flight of thoughts [134, p. 24]. In spite of the obvious vagueness of this hypothesis its trueness has been confirmed in fMRI experiments. It has been found that in drawing test cerebellum is activated only in cases, when theme of drawing demands creative efforts from participants [135].

10. PROFESSIONAL PREDICTIONS FOR HUMAN-LEVEL AI

Andrew Ng makes very skeptical predictions: "Maybe in hundreds of years, technology will advance to a point where there could be a chance of evil killer robots" [136]. "May be hundreds of years from now, may be thousands of years from now—I don't know—may be there will be some AI that turn evil" [3].

Geoffrey Hinton makes a moderate prediction: "I refuse to say anything beyond five years because I don't think we can see much beyond five years" [137].

Shane Legg, DeepMind cofounder, ~~used to make~~ predictions about AGI at the end of ~~each year, here is the last one~~ [138]: "I give it a log-normal distribution with a mean of 2028 and a mode of 2025, under the assumption that nothing crazy happens like a nuclear war. I'd also like to add to this prediction that I expect to see an impressive proto-AGI within the next 8 years". Figure 2 shows the predicted log-normal distribution.

This prediction has been made at the end of 2011. However, it's widely held belief that progress in AI was somewhat unpredictably great after 2011 so it's very reasonable to expect that predictions didn't become to be more pessimistic. In recent 5 years there were perhaps even much more than one revolution in AI field, so it seems quite reasonable that another 5 years can make another very big difference.

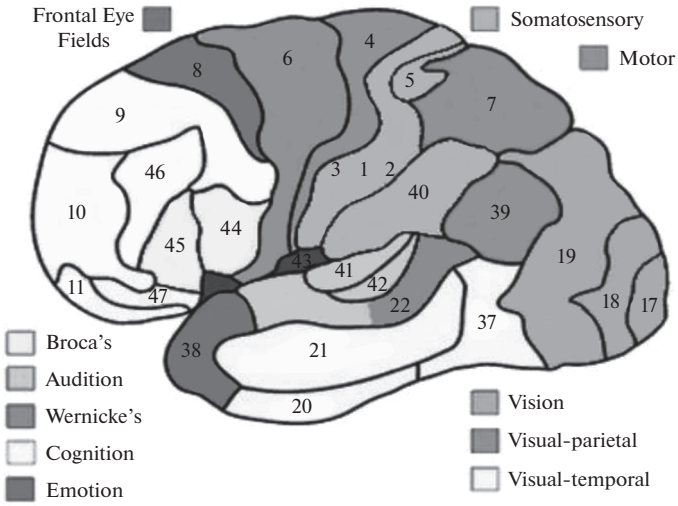

Fig. 3. Human brain, lateral view [165].

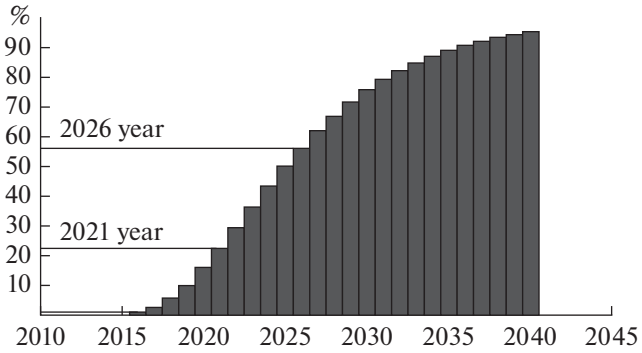

Fig. 4. Shane Legg predictions [165].

For more predictions, see [139]. Here, we illustrated each viewpoint with the quotation of just one scientist. For each of these viewpoints, there are many AI researchers supporting it [139]. A survey of expert opinions on the future of AI in 2012/2013 [140] might also be interesting. However, most of involved people there (even among "top 100" group) are not very much involved with ongoing deep learning progress. Nevertheless the predictions they make are also not too pessimistic. For this article it's quite enough that we are seriously unsure about probabilities of human level AI in next ten years.

Also, it should be mentioned the early prediction, claimed by David Marr Memorial Brain Reverse Engineering Laboratory (Project was active in 2011–2015, Project Manager Witali L. Dunin-Barkowski) in 2011 was: "... It will be possible ... to make (in yrs. 2018–2020) a full-scale working analog of the human brain..." [141].

11. WOULD GENERAL (STRONG) AI BE BENEFICIAL OR DANGEROUS FOR HUMANS?

11.1. Optimistic Arguments

The promising approach to solve/alleviate AI safety problem is in use of deep learning to teach AI our human values given some dataset of ethical problems. Given sufficiently big and broad dataset, we can get sufficiently friendly AI, at least friendlier than most humans can be. However this approach doesn't solve all problems of AI safety [142]. Apart from that we can use inverse reinforcement learning but still would need the benevolence/malevolence dataset of ethical problems to test AIs. This approach has been initially formulated in "The maverick nanny with a dopamine drip" [143] but it's arguably better formulated here [142, 144, 145].

We can challenge AI to rank several solutions (already written by dataset creators) for problems in benevolence/malevolence dataset. It would be easy and fast to check. However, we also want to check the ability of AI to propose its own solutions. This is much harder problem because it requires humans for evaluation. So in intermediate steps we get something like the CEV [146] of Amazon Mechanical Turk. The final version would be checked not by AMT but by a worldwide community including politicians, scientists etc. The very process of checking may last months if not years especially if considering inevitable hot discussions about controversial examples in the dataset. Meantime, people who don't care too much about AI safety would be able to launch their unsafe AGIs.

What would people include in the benevolence/malevolence dataset? Most probably people want some scientific research from AI: a cure for cancer, cold thermonuclear synthesis, AI safety, etc. The dataset would certainly include many examples teaching AI to consult with humans on any serious actions which AI might want to follow and tell humans instantly any insights which arise in AI. Also, MIRI/FHI have hundreds of good papers which might be used to create examples for such a dataset [147].

11.2. Pessimistic Arguments

Most of MIRI research on this topic [147, 148] comes to warning conclusions. The best easy introduction is [2] (or more popular [149] and [150]) though for serious understanding of MIRI arguments, "Superintelligence" [147] and "AI-Foom debate" [2] is very much required. Here we provide our own understanding and review of MIRI arguments which might differ from their official views but still is strongly based on them.

First, we note that it is very profitable to give your AI maximal abilities and permissions.

Those corporations would win their markets which give their AIs direct unrestricted internet access to advertise their products, to collect user's feedback, to build positive impression about company and negative one about competitors, to research users' behavior etc. Those companies would win which use their AIs to invent quantum computing (AI is already used by quantum computer scientists to choose experiments [151, 152]), which allow AI to improve it's own algorithms (including quantum implementation), and even to invent thermonuclear synthesis, asteroid mining etc. All arguments in this paragraph are also true for countries and their defense departments.

Then, human-level AI might quickly become very superhuman-level AI. According to Andrew Karpathy: "I consider chimp-level AI to be equally scary, because going from chimp to humans took nature only a blink of an eye on evolutionary time scales, and I suspect that might be the case in our own work as well. Similarly, my feeling is that once we get to that level it will be easy to overshoot and get to superintelligence" [153].

It takes years or decades for us people to teach our knowledge to other people while AI can almost instantly create full working copies of itself by simple copying. It has a unique memory. A student forgets all the stuff very quickly after exams. AI just saves its configuration just before exams and loads it again when it's needed to solve similar tasks. Modern CNNs do not only recognize images better than human but also make it several orders of magnitude faster. The same holds true for LSTMs in translation, natural language generation etc. Given all aforementioned advantages, AI would quickly learn all literature and video courses on psychology, would chat with thousands of people simultaneously and so would become great psychologist. For same reasons, it would become great scientist, great poet, great businessman, great politician, etc. It would be able to easily manipulate people and lead peoples.

Third fatal but profitable error: direct internet access. If someone gives internet access to human level AI, it would be able to hack millions of computers and run it's own copies or subagents on them like [154]. After that it might earn billion dollars in internet. It would be able to hire anonymously thousands of people to make or buy dexterous robots, 3D-printers, biological labs and even a space rocket. AI would write great clever software to control it's robots.

Fourth, superhuman-level AI has ability to get ultimate power over the Earth. There is a simple yet effective baseline solution. AI might create a combination of lethal viruses and bacteria or some other weapon of mass destruction to be able to kill every human on Earth. You can't guarantee efficient control over something that is much smarter than you. After all, several people almost took over the world, so why superhuman AI can not?

After that, it would destroy humanity, likely as a side-effect.

What would do AI to us if it has full power on Earth? If it's indifferent to us then it will eliminate us as a side effect. It's just what does indifference mean when you are dealing with unbelievably powerful creature solving it's own problems using power of the local Dyson sphere. However if it's not indifferent to us

then everything might be even worse. If it likes us it might decide to insert electrodes in our brains giving us the utmost pleasure but no motivation of doing something. The very opposite would be true if it dislikes us or if it was partially hardcoded to like us but that hardcoding contained a hard-to-catch mistake. Mistakes are almost inevitable when trying to partially hardcode something which is many orders of magnitude smarter and more powerful than you.

There is nothing but vague intuition behind thought that superhuman-level AI would take care of us. There are no physical laws for it to have a special interest in people, in sharing oil/fields/etc. resources with us. Even if AI would care about us, it's very questionable if that care would be appropriate from our today moral standards. We certainly shouldn't take that for granted and to risk everything we have. The fate of humanity would depend on decisions of strong AI once and forever.

A lot of other pessimistic scenarios has been proposed and discussed at length [155–159]. The problem has been discussed for years with no reasonable approach really proposed up to the present moment.

12. CONCLUSIONS

There are many good arguments arguing that human level AI would be constructed during next 5 to 10 years. We are aware of contrary opinions but none of them is explicitly based on thorough analysis of current trends in deep learning and brain reverse engineering. They are in fact based on experience of prominent AI scientists, world acknowledged authoritative scholars, who claim such opinions. However there are other true professionals, AI scientists, who claim that human level AI is quite probable much sooner—in next 5 to 10 years, or less. So it's a good idea to work in the field being ready to any scenario in the field, to their advantages and as well as to their disadvantages.

APPENDIX. ESTIMATION OF NEURAL NETWORK EQUIVALENT AGE

In [36], the age of their model was estimated as 4.45 years. It answered 54.06% questions correctly. They investigate for many questions the youngest age group that could answer it. The results are: age 3–4, 15.3%; age 5–8, 39.7%; age 9–12, 28.4%; age 13–17, 11.2%; age 18+, 5.5%. The sum equals 100% however 18+ humans answer 83.3% correctly. We could estimate that 8 year model must answer understandable for age 8 15.3 + 39.7% = 55% questions with 83.3% accuracy and other 45% with accuracy equal to baseline model. If they select the most popular answer for each question type, they get 36.18%. It's reasonable to estimate the age of that baseline model as 0 years because it's not real knowledge but it depends on this distinct dataset statistics. However, there are two baseline models in article, with 36.18 and 40.61% accuracy, both are quite simple. Which one to choose? Let's calculate. Let's define accuracy of 8 year old model as y, accuracy of 4 year old model as x, baseline accuracy as t. We have following equations: $4 + 4(54.06 - x)/(y - x) = 4.45$ $55 \times 0.833 + 45t = y$ $15.3 \times 0.833 + 84.7t = x$ with solution: $t = 46.86\%$, $x = 52.4\%$, $y = 66.9\%$. So we estimate age of model with 57.6% accuracy as $4 + 4(57.6 - 52.4)/(66.9 - 52.4) = 5.45$ years. We estimate age of model with 60.4% accuracy as $4 + 4(60.4 - 52.4)/(66.9 - 52.4) = 6.2$ years.

ACKNOWLEDGMENTS

This work was supported by RFBR Grant 16-07-01059 and by the National Technological Initiative (Russian Federation) project "Artificial Neural Intelligence iPavlov".

REFERENCES

1. https://www.frontiersin.org/research-topics/6714/toward-and-beyond-human-level-ai.
2. Yudkowsky, E., Artificial Intelligence As a Positive and Negative Factor in Global Risk. https://intelligence.org/files/AIPosNegFactor.pdf.
3. Muerlhauser, L., What is AGI? https//intelligence.org/2013/08/11/what-is-agi/.
4. Grace, K., Algorithmic Progress in Six Domains. https://intelligence.org/files/AlgorithmicProgress.pdf.
5. Schmidhuber, J., Deep Learning in Neural Networks: An Overview. http://arxiv.org/abs/1404.7828.
6. Dunin-Barkowski, W. and Solovyeva, K., Pavlov Principle in Brain Reverse Engineering, *Neuroinformatics 2016, Proceedings of the XVIII International Conference on Neuroinformatics*, Part 1, Moscow: MEPHI, 2016.
7. Pavlov, I., Conditioned reflexes, in *Twenty Years of Experience of Objective Studies of Higher Nervous Activity (Behavior) of Animals by I.P. Pavlov*, 10th ed. (1st ed. in 1923), Moscow: Nauka, 1973, pp. 485–502 [in Russian].
8. Gatys, L.A., Ecker, A.S., and Bethge, M., A Neural Algorithm of Artistic Style. http://arxiv.org/abs/1508.06576.

9. Shannon, C.E., Prediction and Entropy of Written English. http://languagelog.ldc.upenn.edu/myl/Shannon1950.pdf.
10. http://cs.fit.edu/~mmahoney/dissertation/entropy1.html.
11. Ghosh, S., Vinyals, O., Strope, B., Roy, S., Dean, T., and Heck, L., Contextual LSTM (CLSTM) Models for Large Scale NLP Tasks. http://arxiv.org/abs/1602.06291.
12. Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., and Wu, Y., Exploring the Limits of Language Modeling. http://arxiv.org/abs/1602.02410.
13. Husz'ar, F., How (not) to Train Your Generative Model: Scheduled Sampling, Likelihood, Adversary? http://arxiv.org/abs/1511.05101.
14. Theis, L., van den Oord, A., and Bethge, M., A Note on the Evaluation of Generative Models. http://arxiv.org/abs/1511.01844.
15. Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A.M., Jozefowicz, R., and Bengio, S., Generating Sentences from a Continuous Space. http://arxiv.org/abs/1511.06349.
16. Arjovsky, M., Chintala, S., and Bottou, L., Wasserstein GAN. https://arxiv.org/pdf/1701.07875.pdf.
17. https://www.cs.sfu.ca/~anoop/papers/pdf/jhu-ws03-report.pdf.
18. Shen, S., Cheng, Y., He, Z., He, W., Wu, H., Sun, M., and Liu, Y., Minimum Risk Training for Neural Machine Translation. http://arxiv.org/abs/1512.02433.
19. Bahdanau, D., Cho, K., and Bengio, Y., Neural Machine Translation by Jointly Learning to Align and Translate. http://arxiv.org/abs/1409.0473.
20. Sennrich, R., Haddow, B., and Birch, A., Improving Neural Machine Translation Models with Monolingual Data. http://arxiv.org/abs/1511.06709.
21. Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., and Makhoul, J., Fast and Robust Neural Network Joint Models for Statistical Machine Translation. http://acl2014.org/acl2014/P14-1/pdf/P14-1129.pdf.
22. Lample, G., Denoyer, L., and Ranzato, M.A., Unsupervised Machine Translation Using Monolingual Corpora Only. http://arXiv: 1711.00043v1.
23. He, K., Zhang, X., Ren, S., and Sun, J., Identity Mappings in Deep Residual Networks. http://arxiv.org/abs/1603.05027.
24. http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-onimagenet/.
25. He, K., Zhang, X., Ren, S., and Sun, J., Deep Residual Learning for Image Recognition. http://arxiv.org/abs/1512.03385.
26. Szegedy, C., Ioffe, S., and Vanhoucke, V., Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. http://arxiv.org/abs/1602.07261.
27. Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K., Deep Networks with Stochastic Depth. http://arxiv.org/abs/1603.09382.
28. http://myungjun-youn-demo.readthedocs.org/en/latest/tutorial/imagenet_full.html.
29. Kokkinos, I., Pushing the Boundaries of Boundary Detection Using Deep Learning. http://arxiv.org/abs/1511.07386.
30. http://vision.stanford.edu/pdf/karpathy14.pdf.
31. Ng, J.Y.-H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., and Toderici, G., Beyond Short Snippets: Deep Networks for Video Classification. http://arxiv. org/abs/1503.08909.
32. http://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakthrough-year-inartificial-intelligence.
33. https://github.com/soumith/convnet-benchmarks.
34. Han, S., Liu, X., Mao, H., Pu, J., Pedram, A., Horowitz, M.A., and Dally, W.J., EIE: Efficient Inference Engine on Compressed Deep Neural Network. http://arxiv.org/abs/1602.01528.
35. Xiong, C., Merity, S., and Socher, R., Dynamic Memory Networks for Visual and Textual Question Answering. http://arxiv.org/abs/1603.01417.
36. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C.L., and Parikh, D., VQA: Visual Question Answering. http://arxiv.org/abs/1505.00468.
37. Yang, Z., He, X., Gao, J., Deng, L., and Smola, A., Stacked Attention Networks for Image Question Answering. http://arxiv.org/abs/1511.02274.
38. http://competitions.codalab.org/competitions/3221#results.
39. https://github.com/samim23/NeuralTalkAnimator.
40. Pan, P., Xu, Z., Yang, Y., Wu, F., and Zhuang, Y., Hierarchical Recurrent Neural Encoder for Video Representation with Application to Captioning. http://arxiv.org/abs/1511. 03476.
41. Yu, H., Wang, J., Huang, Z., Yang, Y., and Xu, W., Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks. http://arxiv.org/abs/1510.07712.
42. Mansimov, E., Parisotto, E., Ba, J.L., and Salakhutdinov, R., Generating Images from Captions with Attention. http://arxiv.org/abs/1511.02793.

43. Yan, X., Yang, J., Sohn, K., and Lee, H., Attribute2Image: Conditional Image Generation from Visual Attribute. http://arxiv.org/abs/1512.00570.
44. Pandey, G. and Dukkipati, A., Variational methods for Conditional Multimodal Learning: Generating Human Faces from Attributes. http://arxiv.org/abs/1603.01801.
45. http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-onlyan-8-word-error-rate/.
46. Wu, Z., Jiang, Y.-G., Wang, X., Ye, H., Xue, X., and Wang, J., Fusing Multi-Stream Deep Networks for Video Classification. http://arxiv.org/abs/1509.06086.
47. Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., and Yuille, A., Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN). http://arxiv.org/abs/1412.6632.
48. Kahou, S.E., Bouthillier, X., Lamblin, L., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., Ferrari, R.C., Mirza, M., Warde-Farley, D., Courville, A., Vincent, P., Memisevic, R., Pal, C., and Bengio, Y., EmoNets: Multimodal Deep Learning Approaches for Emotion Recognition in Video. http://arxiv.org/abs/1503.01800.
49. Moon, S., Kim, S., and Wang, H., Multimodal Transfer Deep Learning with Applications in Audio-Visual Recognition. http://arxiv.org/abs/1412.3121.
50. Rohrbach, A., Rohrbach, M., Hu, R., Darrell, T., and Schiele, B., Grounding of Textual Phrases in Images by Reconstruction. http://arxiv.org/abs/1511.03745.
51. Yang, Y., Li, Y., Fermuller, C., and Aloimonos, Y., Neural Self Talk: Image Understanding via Continuous Questioning and Answering. http://arxiv.org/abs/1512.03460.
52. Silver, D., Schrittwieser, J., et al., Mastering the game of Go without human knowledge, *Nature*, 2017, vol. 550, pp. 354–359.
53. Gu, S., Lillicrap, T.P., Sutskever, I., and Levine, S., Continuous Deep Q-Learning with Model-Based Acceleration. http://arxiv.org/abs/1603.00748.
54. Schulman, J., Moritz, P., Levine, S., Jordan, M., and Abbeel, P., High-Dimensional Continuous Control Using Generalized Advantage Estimation. http://arxiv.org/abs/1506.02438.
55. Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D., Continuous Control with Deep Reinforcement Learning. http://arxiv.org/abs/1509.02971.
56. Finn, C., Levine, S., and Abbeel, P., Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization. http://arxiv.org/abs/1603.00448.
57. Hausknecht, M. and Stone, P., Deep Reinforcement Learning in Parameterized Action Space. http://arxiv.org/abs/1511.04143.
58. Heess N., Wayne, G., Silver, D., Lillicrap, T., Tassa, Y., and Erez, T., Learning Continuous Control Policies by Stochastic Value Gradients. http://arxiv.org/abs/1510.09142.
59. Heess, N., Hunt, J.J., Lillicrap, T.P., and Silver, D., Memory-Based Control with Recurrent Neural Networks. http://arxiv.org/abs/1512.04455.
60. Balduzzi, D. and Ghifary, M., Compatible Value Gradients for Reinforcement Learning of Continuous Deep Policies. http://arxiv.org/abs/1509.03005.
61. Dulac-Arnold, G., Evans, R., Sunehag, P., and Coppin, B., Reinforcement Learning in Large Discrete Action Spaces. http://arxiv.org/abs/1512.07679.
62. http://togelius.blogspot.ru/2016/01/why-video-games-are-essential-for.html.
63. He, J., Chen, J., He, X., Gao, J., Li, L., Deng, L., and Ostendorf, M., Deep Reinforcement Learning with an Action Space Defined by Natural Language. http://arxiv.org/ abs/1511.04636.
64. Narasimhan, K., Kulkarni, T., and Barzilay, R., Language Understanding for Text-Based Games Using Deep Reinforcement Learning. http://arxiv.org/abs/1506.08941.
65. Sutton, R.S. and Barto, A.G., *Reinforcement Learning: An Introduction*, Massachusetts: MIT Press, 2018.
66. Dream. https://en.wikipedia.org/wiki/Dream.
67. Radford, A., Metz, L., and Chintala, S., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. http://arxiv.org/abs/1511.06434.
68. https://www.facebook.com/yann.lecun/posts/10153269667222143.
69. Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R.S., Torralba, A., Urtasun, R., and Fidler, S., Skip-Thought Vectors. http://arxiv.org/abs/1506.06726.
70. Kingma, D.P. and Welling, M., Auto-Encoding Variational Bayes. http://arxiv.org/abs/ 1312.6114.
71. Rezende, D.J., Mohamed, S., and Wierstra, D., Stochastic Backpropagation and Approximate Inference in Deep Generative Models. http://arxiv.org/abs/1401.4082.
72. Eslami, S.M.A., Heess, N., Weber, T., Tassa, Y., Kavukcuoglu, K., and Hinton, G.E., Attend, Infer, Repeat: Fast Scene Understanding with Generative Models. http: //arxiv.org/abs/1603.08575.
73. https://en.wikipedia.org/wiki/Tetra-amelia_syndrome.
74. https://en.wikipedia.org/wiki/Hirotada_Ototake.

75. https://en.wikipedia.org/wiki/Nick_Vujicic.
76. https://en.wikipedia.org/wiki/Prince_Randian.
77. http://googleresearch.blogspot.ru/2016/03/deep-learning-for-robots-learning-from.html.
78. https://kaiserfamilyfoundation.files.wordpress.com/2013/04/8010.pdf.
79. http://bbc.com/news/technology-28677674.
80. https://faculty.washington.edu/chudler/facts.html.
81. Kennedy, D.N., Lange N., et al., Gyri of the human neocortex, *Cereb. Cortex*, 1998, vol. 8, pp. 372–384.
82. Maruoka, H., Nakagawa, N., Tsuruno, S., Sakai, S., Yoneda, T., and Hosoya, T., Lattice system of functionally distinct cell types in the neocortex, *Science*, 2017, vol. 358, pp. 610–615.
83. Vinyals, O. and Le, Q., A Neural Conversational Model. http://arxiv.org/abs/1506.05869.
84. Ghosh, S., Vinyals, O., Strope, B., Roy, S., Dean, T., and Heck, L., Contextual LSTM (CLSTM) Models for Large Scale NLP Tasks. http://arxiv.org/abs/1602.06291.
85. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M., Playing Atari with Deep Reinforcement Learning. http://arxiv.org/abs/1312.5602.
86. Lake, B.M., Ullman, T.D., Tenenbaum, J.B., and Gersshman, S.J., Building machines that learn and think like people, *Behav. Brain Sci.*, 2017, vol. 40, p. 72. doi doi 10.1017/S0140525X16001837
87. Eberle, A.L., Mikula, S., Schalek, R., Lichtman, J., Knothe Tate, M.L., and Zeidler, D., High-Resolution, High-Throughput Imaging with a Multibeam Scanning Electron Microscope. http://onlinelibrary.wiley.com/doi/10.1111/jmi.12224/pdf.
88. http://news.harvard.edu/gazette/story/2016/01/28m-challenge-figure-out-why-brains-are-so-goodat-learning/.
89. Kasthuri, N., Hayworth, K.J., et al., Saturated Reconstruction of a Volume of Neocortex. https://www.mcb.harvard.edu/mcb_files/media/editor_uploads/2015/07/PIIS0092867415008247.pdf.
90. Lillicrap, T.P., Cownden, D., Tweed, D.B., and Akerman, C.J., Random Feedback Weights Support Learning in Deep Neural Networks. http://arxiv.org/abs/1411.0247.
91. Lillicrap, T.P., Cownden, D., Tweed, D.B., and Akerman, C.J., Random synaptic feedback weights support error backpropagation for deep learning, *Nat. Commun.*, 2016, vol. 7, no. 13276, p. 10.
92. Hebb, D., *The Organization of Behavior*, New York: Wiley, 1949.
93. Noklund, A., Direct Feedback Alignment Provides Learning in Deep Neural Networks. https://arxiv.org/pdf/1609.01596.pdf.
94. Karandashev, I.M. and Dunin-Barkowski, W.L., Computational verification of approximate probabilistic estimates of operational efficiency of random neural networks, *Opt. Mem. Neural Networks*, 2015, vol. 24, pp. 8–17.
95. Solovyeva, K.P., Karandashev, I.M., Zhavoronkov, A., and Dunin-Barkowski, W.L., Models of innate neural attractors and their applications for neural information processing, *Front. Syst. Neurosci.*, 2016, vol. 9, no. 176, p. 13.
96. Gilmer, J., Raffel, C., Schoenholz, S.S., Ragnu, M., and Sohl-Dickstein, J., *Explaining the Learning Dynamics of Direct Feedback Alignment*, *Workshop Track – ICLR* 2017, p. 4.
97. Bengio, Y. and Fische, A., Early Inference in Energy-Based Models Approximates BackPropagation. http://arxiv.org/abs/1510.02777.
98. Baldi, P. and Sadowski, P., The Ebb and Flow of Deep Learning: a Theory of Local Learning. http://arxiv.org/abs/1506.06472.
99. Bengio, Y., Lee, D.-H., Bornschein, J., and Lin, Z., Towards Biologically Plausible Deep Learning. http://arxiv.org/abs/1502.04156.
100. Liao, Q., Leibo, J.Z., and Poggio, T., How Important is Weight Symmetry in Backpropagation? http://arxiv.org/abs/1510.05067.
101. Bengio, Y., Mesnard, T., Fischer, A., Zhang, S., and Wu, Y., STDP As Presynaptic Activity Times Rate of Change of Postsynaptic Activity. http://arxiv.org/abs/1509.05936.
102. Ollivier, Y., Tallec, C., and Charpiat, G., Training Recurrent Networks Online without Backtracking. http://arxiv.org/abs/1507.07680.
103. Andrychowicz, M. and Kurach, K., Learning Efficient Algorithms with Hierarchical Attentive Memory. http://arxiv.org/abs/1602.03218.
104. Zaremba, W. and Sutskever, I., Reinforcement Learning Neural Turing Machines. http://arxiv.org/abs/1505.00521.
105. Neelakantan, A., Le, Q.V., and Sutskever, I., Neural Programmer: Inducing Latent Programs with Gradient Descent. http://arxiv.org/abs/1511.04834.
106. Fu, J., Luo, H., Feng, J., Low, K.H., and Chua, T.-S., DrMAD: Distilling ReverseMode Automatic Differentiation for Optimizing Hyperparameters of Deep Neural Networks. arxiv. org/abs/1601.00917.
107. Loshchilov, I. and Hutter, F., Online Batch Selection for Faster Training of Neural Networks. http://arxiv.org/abs/1511.06343.

108. Bengio, E., Bacon, P.-L., Pineau, J., and Precup, V., Conditional Computation in Neural Networks for faster models. http://arxiv.org/abs/1511.06297.
109. Shah, V., Asteris, M., Kyrillidis, A., and Sanghavi, S., Trading-Off Variance and Complexity in Stochastic Gradient Descent. http://arxiv.org/abs/1603.06861.
110. Arjovsky, M., Shah, A., and Bengio, Y., Unitary Evolution Recurrent Neural Networks. http://arxiv.org/abs/1511.06464.
111. Novikov, A., Podoprikhin, D., Osokin, A., and Vetrov, D., Tensorizing Neural Networks. http://arxiv.org/abs/1509.06569.
112. Rhu, M., Gimelshein, N., Clemons, J., Zulfiqar, A., and Keckler, S.V., Virtualizing Deep Neural Networks for Memory-Efficient Neural Network Design. http://arxiv.org/abs/1602.08124.
113. Chen, T., Goodfellow, I., and Shlens, J., Net2Net: Accelerating Learning via Knowledge Transfer. http://arxiv.org/abs/1511.05641.
114. Wei, T., Wang, C., Rui, Y., and Chen, C.W., Network Morphism. http://arxiv.org/abs/1603.01670.
115. https://github.com/BVLC/caffe/wiki/Model-Zoo.
116. http://myungjun-youn-demo.readthedocs.org/en/latest/pretrained.html.
117. http://www.vlfeat.org/matconvnet/pretrained/.
118. Pascanu, R., Gulcehre, C., Cho, K., and Bengio, Y., How to Construct Deep Recurrent Neural Networks. http://arxiv.org/abs/1312.6026.
119. Zhang, S., Wu, Y., Che, T., Lin, Z., Memisevic, R., Salakhutdinov, R., and Bengio, Y., Architectural Complexity Measures of Recurrent Neural Networks. http://arxiv.org/abs/1602.08210.
120. Cooijmans, T., Ballas, N., Laurent, C., and Courville, A., Recurrent Batch Normalization. http://arxiv.org/abs/1603.09025.
121. Semeniuta, S., Severyn, A., and Barth, E., Recurrent Dropout without Memory Loss. http://arxiv.org/abs/1603.05118.
122. Serban, I.V., Sordoni, A., Bengio, Y., Courville, A., and Pineau, J., Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. http://arxiv.org/abs/1507.04808.
123. Yo, K., Zweig, G., and Peng, B., Attention with Intention for a Neural Network Conversation Model. http://arxiv.org/abs/1510.08565.
124. Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Simonsen, J.G., and Nie, J., A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion. http://arxiv.org/abs/1507.02221.
125. Graves, A., Adaptive Computation Time for Recurrent Neural Networks. http://arxiv.org/ abs/1603.08983.
126. Gokmen, T. and Vlasov, Yu., Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices. http://arxiv.org/abs/1603.07341.
127. Negrov, D., Karandashev, I., Shakirov, V., Matveev, Yu., Dunin-Barkowski, W., and Zenkevich, A., An approximate backpropagation learning rule for memristor-based neural networks using synaptic plasticity, *Neurocomputing*, 2017, vol. 237, pp. 193–199.
128. http://www.technologyreview.com/news/544421/googles-quantum-dream-machine/.
129. Lloyd, S., Mohseni, M., and Patrick, P., Quantum Algorithms for Supervised and Unsupervised Machine Learning. http://arxiv.org/abs/1307.0411.
130. Bernien, H., Schwartz, S., Keesling, A., Levine, H., Omran, A., Pichler, H., Choi, S., Zibrov, A.S., Endres, M., Gremer, M., Vuletic, V., and Lukin, M.D., Probing Many Body Dynamics on a 51-Atom Quantum Simulator. http://arxiv.org/abs/1707.04344.
131. https://twitter.com/andrewyng/status/693182932530262016.
132. Quora, Ten Things Everyone Should Know About Machine Learning, *Forbes*, 2017, September 6.
133. Jastrzebski, S., Le'sniak, D., and Czarnecki, W.M., Learning to SMILE(S). http://arxiv.org/abs/1602.06289.
134. Dunin-Barkowski, W.L., *Theory of cerebellum, Lectures on Neuroinformatics*, Moscow: MEPHI, 2010, pp. 15–48 [in Russian].
135. Saggar, M., Quintin, E.M., et al., Pictionary-based fMRI paradigm to study the neural correlates of spontaneous improvisation and figural creativity, *Sci. Rep.*, 2015, vol. 5: 10894, p. 11.
136. http://blogs.nvidia.com/blog/2015/03/19/riding-the-ai-rocket-top-artificial-intelligence-researchersays-robots-wont-kill-us-all/.
137. http://www.macleans.ca/society/science/the-meaning-of-alphago-the-ai-program-that-beat-a-gochamp/.
138. http://www.vetta.org/2011/12/goodbye-2011-hello-2012/.
139. http://slatestarcodex.com/2015/05/22/ai-researchers-on-ai-risk/.
140. Muller, V.C. and Bostrom, N., Future Progress in Artificial Intelligence: A Survey of Expert Opinion. https://intelligence.org/files/AlgorithmicProgress.pdf.
141. http://rebrain.2045.com.
142. http://www.facebook.com/groups/467062423469736/permalink/573838239458820/.

143. Loosemore, R., The Maverick Nanny with a Dopamine Drip: Debunking Fallacies in the Theory of AI Motivation. http://richardloosemore.com/docs/2014a_MaverickNanny_rpwl.pdf.
144. Loosemore, R., Defining Benevolence in the context of Safe AI. http://ieet.org/index.php/IEET/more/loosemore20141210.
145. http://www.facebook.com/groups/467062423469736/permalink/532404873602157/.
146. Yudkowsky, E., Coherent Extrapolated Volition. https://intelligence.org/files/CEV.pdf.
147. Bostrom, N., *Superintelligence: Pathys, Dangers, Strategies, MP3 CD*, Oxford: Oxford Univ. Press, 2014.
148. https://intelligence.org/our-research/.
149. http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html.
150. http://blog.samaltman.com/machine-intelligence-part-1.
151. August, M. and Ni, X., Using Recurrent Neural Networks to Optimize Dynamical Decoupling for Quantum Memory. http://arxiv.org/abs/1604.00279.
152. http://www.cnet.com/news/quantum-science-is-so-weird-that-ai-is-choosing-the-experiments/.
153. http://singularityhub.com/2015/12/20/inside-openai-will-transparency-protect-us-from-artificialintelligence-run-amok/.
154. http://en.wikipedia.org/wiki/Storm_botnet.
155. http://lesswrong.com/lw/691/qa_with_shane_legg_on_risks_from_ai/.
156. https://backchannel.com/how-elon-musk-and-y-combinator-plan-to-stop-computers-from-takingover-17e0e27dd02a.
157. http://www.nickbostrom.com/superintelligentwill.pdf.
158. de Grey, A. and Rae, M., *Ending Aging: The Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime*, New York: St. Martin's Press, 2008.
159. Yampolsky, R., Taxonomy of Pathways to Dangerous AI. http://arxiv.org/abs/1511.03246.
160. Ji, Y., Cohn, T., Kong, L., Dyer, C., and Eisenstein, J., Document Context Language Models. http://arxiv.org/abs/1511.03962.
161. http://host.robots.ox.ac.uk:8080/leaderboard/displaylb.php?challengeid=11&compid=4.
162. Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., and Ng, A.Y., Deep Speech: Scaling up End-to-End Speech Recognition. http://arxiv.org/abs/1412.5567.
163. Amodei, D., Anubhai, R., et al., Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. http://arxiv.org/abs/1512.02595.
164. https://github.com/hanzhanggit/StackGAN.
165. https://habrahabr.ru/company/parallels/blog/331726/.

SPELL: 1. feedforward, 2. backprop