Machine Intelligence: from Turing Test to Generative Pretrained Transformers

Fethallah Hadjila, Mohammed Merzoug and Amine Belabed

December 27, 2023

Machine intelligence: from Turing test to generative pretrained transformers

Hadjila Fethallah
Computer Science Department, LRIT Laboratory
University of Tlemcen
fethallah.hadjila@univ-tlemcen.dz
Merzoug Mohammed
Computer Science Department, LRIT Laboratory
University of Tlemcen
mohammed.merzoug@univ-tlemcen.dz
Belabed Amine
Computer Science Department, LRIT Laboratory
University of Tlemcen
amine.belabed@univ-tlemcen.dz

***Abstract*—Artificial intelligence (A.I) has impressively evolved during the last seventy years, especially with the advances of machine learning. The objective of this paper is to present a survey on the chronological development of A.I from its early beginning up to the last breakthroughs. Namely, we will cover the Turing test ingredients, dark era of neural networks, resurgence of neural networks through back-propagation, ensemble learning, and deep learning. We also review the actual challenges and the eventual dangers that still face the research community.**

Keywords– History of AI, Machine learning, Deep learning, Convolutional neural networks.

1. Introduction

Understanding and simulating the human brain capabilities is one of the old objectives of the computer science community. It is worth noting that the human intelligence is a delicate term to define and it spawned much debate and ink to flow in the past decades. In 1994, a group of psychology researchers led by Linda Gottfredson proposed a definition of human intelligence that is specified as follows: "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience." [1]. Even if this definition is not a consensual one, we can use it to draft a first definition for the artificial intelligence (A.I) utterance. In this line of thought, A.I is the simulation of the above-mentioned brain capabilities on machines. Consequently, some software systems may not be considered as intelligent (by other researchers) if they cannot fulfill all the capabilities mentioned above. The reasoning skill means the ability to think and compose inference rules and axioms to build new valid knowledge that can be considered as theorems. The planning skill means the ability to produce a sequence of actions that map an initial state of affairs into a goal state. Thinking abstractly is a kind of reasoning or problem resolution that builds solutions or target objects using abstract input blocks, the target built objects must show the desired properties while masking the less important details. The comprehension of complex ideas means the ability to process the natural language (NLP) while performing all necessary steps in terms of acoustic/phonetic analysis, morphological analysis, syntactic analysis, semantic analysis, and pragmatic analysis. The last skill concerns the ability of training and acquiring new knowledge and competencies through the rewiring (connections' update) of the brain, we can distinguish off-line learning (which may consume much time before the effective utilization) and real-time learning which must meet the delays imposed by the current problem.

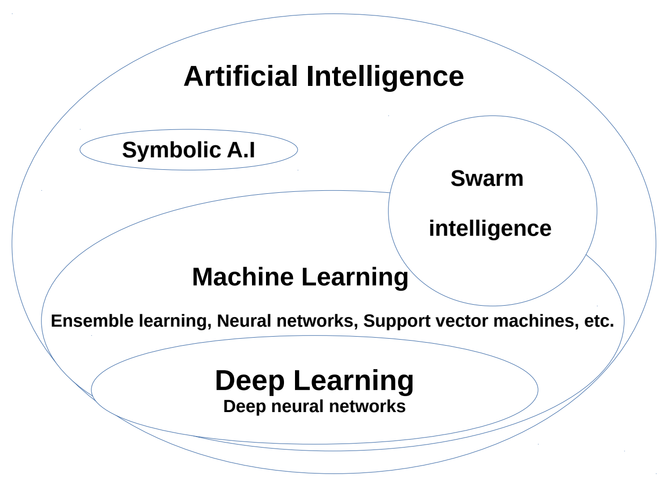

Figure 1. A.I ingredients.

In Figure 1, we point out that A.I is mainly constituted of symbolic intelligence, machine learning, and emergentist intelligence (i.e., swarm and connectionist intelligence). This entire AI may pass different levels of perfection. We distinguish narrow AI (or specialised /weak A.I), Strong A.I sometimes called Artificial general intelligence (i.e., AGI) and finally Super A.I. Narrow AI is actual state of machine intelligence. It mainly means that the developed (smart) softwares are better than a human in a specific task and a specific context. For instance, in the object recognition field, we witnessed many models (starting from 2015) that bypass the human capabilities (see Figure 4). According to Ray Kurzweil [2], the term "narrow AI" refers to the creation of systems that carry out specific "intelligent" behaviors in specific contexts. As concrete Narrow AI systems, we can cite object recommender systems, self-driving vehicles, Chatbots, and IBM Watson.

We can find other artificial systems that are way better than human being in a given task, but this cannot be generalised to all tasks. The bottom line is that narrow A.I is better than humans in a single task (or related tasks in the extreme case), but the human is better than narrow A.I when we consider all tasks. In contrast to that, AGI or strong A.I is a kind of software systems that have human like capabilities in a variety of tasks and contexts. Their generalisation capability is as good as that of humans, and they may surpass them in some tasks. Goertzel describes AGI as follows: "General intelligence involves the ability to achieve a variety of goals, and carry out a variety of tasks, in a variety of different contexts and environments" [3]. These tasks include the skills found in the Turing test and they involve perception, reasoning, planning, real-time /batch learning, and natural language processing. Until now, we have not such an existing AGI system; researchers estimate that the future versions of GPTs (Generative pre-trained Transformers) may have these capabilities or at least a percentage of them around 2050. Artificial Super Intelligence ASI is the last level of A.I. In this case, ASI is better than human in every task and every context. It also enjoys the consciousness capability and other kinds of feelings. Certainly ASI will constitute a real danger for the other species, including humanity.

To meet the aforementioned capabilities A.I has gone through a remarkable path and has witnessed many ups and downs. The A.I history begins with the works of Warren McCulloch and Walter Pitts in 1943 that created the first artificial neuron. Then, the Turing test specified the main ingredient of A.I in 1950, and in this perspective, we witnessed the effective birth of A.I with the Dartmouth Workshop in 1956. This conference laid the foundation for symbolic A.I and logic-based reasoning which flourished in the 1970s and 1980s. Almost in the same period (1960s-1970s), the dark era of machine learning (or connectionist A.I winter) is experienced with the book of Minsky and Papert, this latter one highlighted the main drawbacks of the Perceptron neural network and encouraged the abandonment of works in this field. The resurgence of neural networks and connectionism in the mid 1980s constituted a pivotal point in the A.I life-cycle and led to the modern era of data-driven machine learning; more specifically, the period of the 1990s and 2000s witnessed the emergence of effective algorithms like multi layer perceptrons [4], support vector machines [5], and decision trees [6]. Nonetheless, the true renaissance of A.I unfolded after the 2010s, as A.I delved into the world of deep learning. Many remarkable frameworks of deep neural networks, such as convolutional neural networks (CNNs) [7], [8], recurrent neural networks (RNNs), [9], [10] and transformers [11], have revolutionized critical fields like computer vision, speech recognition, and natural language processing. Nowadays A.I systems are providing brilliant solutions that transform a multitude of sectors (including healthcare, industry, and cities) into a highly smart ecosystem that eases and promotes the life of human beings. In this work, we mainly discuss the chronological development of A.I while showing the most eminent successes of A.I softwares. We also review some challenges, as well as some threats that must be handled in order avoid the eventual deviation of A.I systems and other potential misuses.

The rest of the paper is structured as follows, in Section 2, we review the main stages through which A.I has gone. In Section 3, we show some prominent successes of A.I systems. In Section 4, we highlight the challenges and threats of A.I. In the last Section, we give a summary about the future applications of A.I.

2. A.I chronological progress

The AGI level of A.I can be realized using various architectures, according to Duch's overview [12], there are three types of approaches (or paradigms) to implement machine intelligence, we have the symbolic way , emergentist way and hybrid way (this view is approximately depicted in Figure 1). Even though this categorization is not a consensual one, it brings an initial view about the progress path of A.I. In what follows, we will review the main stages through which A.I has gone.

2.1. Early foundations (1950s-1960s)

In 1950, Turing proposed one of the most accepted definitions for A.I, the main idea consists of performing a test in which there are three entities that communicate through a network (see Figure 2):

- **Human interrogator:** his role is to ask a series of questions written in natural language to both interlocutors (the human and the machine). The interrogator does not know the exact the location of both interlocutors and must only use the answers to perform the distinction.
- **Human (interlocutor):** his role is to answer the questions via the network and using written natural language.
- **Machine (interlocutor):** its role is to answer the questions via the network and using written natural language.

If the interrogator can not distinguish between the two interlocutors after the end of questions' session, then the machine is qualified to be intelligent. To succeed this test, the machine must be endowed with various kill including natural language processing, knowledge representation and reasoning, learning from examples. Moreover, the advanced version of the Turing test must also include computer vision in order to process multimedia data, and also robotic parts to perform real world actions and planning.

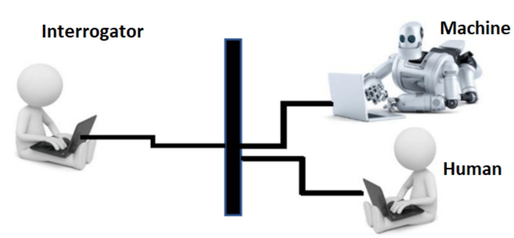

Figure 2. Turing test.

2.2. A.I Dark Era and Expert Systems (1960s-1980s).

In 1956, a group of researchers organized the first conference on A.I in Dartmouth and this latter one opened the door for the advances in the field of symbolic A.I. In particular, many theorem provers and general problem solvers have been developed. Moreover, the formalization of the human expertise in terms of rules and other knowledge formalisms led to the creation of rule-based systems and knowledge-based systems. As an example, we cite the MYCIN system which is developed in the early 1970s. MYCIN is a rule-based system that uses backward chaining in order to deduce the bacteria that are involved in many infection diseases including septicemia and meningitis. In the same period (1960-1980), the ”machine learning winter” entered the scene with the work of Minsky and Papert in the late 1960s. According to these authors, the connectionist approaches (Perceptrons) are very limited and can't even learn simple functions such as XOR. Due to this work, the research in this field had been frozen for at least fifteen years.

2.3. Connectionism, neural networks, and statistical learning theory (1980s-1990s)

In 1986, Geoffrey Hinton and his colleagues developed a method called ”back-propagation” to train multilayer neural networks, and this allows the learning of complex mathematical functions and ensures the ultimate objective of universal approximation. On the other hand, an innovative learning theory was proposed in 1984 by Leslie Valiant in order to mathematically analyse machine learning algorithms and models. This theory (also called probably-approximately correct (PAC) learning) [13] specifies the necessary conditions that help building learning models with a good generalization capability. Starting from this theory, researchers have built many learning models with guaranteed performance (and generalization power) including support vector machines and ensemble learning methods. It is worth noting that this type of machine learning models have realized a spectacular success on small size and medium size problems such as handwritten letters and digits recognition.

2.4. The Rise of machine learning (1990s-2000s)

This era is mainly marked by the creation and development of models that respect the PAC learning paradigm and mainly the successful development of support vector machines and its upgrades in the 1990s. The genesis of ensemble learning in the late 1990s and the early 2000s added a powerful layer in machine learning and reinforced the PAC learning paradigm. Both the boosting family which essentially represented by Adaboost [14] and the bagging family which is predominantly represented by Random forest [15] are able to convert a set of simple learning models (called weak learners) into a strong learner which satisfies the PAC learning objectives.

2.5. Deep learning revolution (2010-Present)

With advent of complex tasks such as perception challenges (in which tremendous and disperate types of data must be effectively handled by A.I models), the inefficiency of current machine learning methods such as SVM [5], hidden markovian models (HMM) [16], and multilayer perceptrons (MLP) [4] was clear, especially in computer vision. For instance, in object recognition, the best models, which preceded the deep learning era, obtained only an error rate equal to 25.8% on the ImageNet challenge [17][1]. In this perspective, Alex krisevski [8] proposed an effective Convolutional neural network (CNN) on ImageNet dataset and decreased the error rate to 16%. This spectacular progress invited all IT companies and research laboratories to turn their efforts to deep neural networks in all perception tasks. In the perspective of CNN, we mainly cite the following advanced architectures: inception networks [18], residual networks [19], and efficient networks [20] for object recognition; U-nets [21] and deeplab [22] for image segmentation. In the subsequent years, many research teams proposed their own deep neural networks in all supervised and unsupervised challenges, and each time a sufficient number of examples is available for a given issue (with an adequate computing power), there will be a successful performance that opens the doors for real breakthroughs. In the context of recurrent networks, we highlight the revival of long short term memory and its competitor Gated recurrent units in crucial tasks such as auto-completion and automatic translation. In 2015, the mechanism of attention [23] was introduced in sequence to sequence models (namely LSTM) to enhance the capabilities of translators and gain better results in handling longer sentences. Thereafter (in 2016), the self-attention [24] concept was coined for single encoder networks in order to learn the context of each word of sentences. Afterwards (in 2017), the self-attention mechanism, residual blocks, and positional encoding were combined to create an innovative sequence to sequence model that impressively succeed in automatic translation. This new architecture was termed Transformer [11] and gained a lot of success in sequential data processing (see Figure 3).

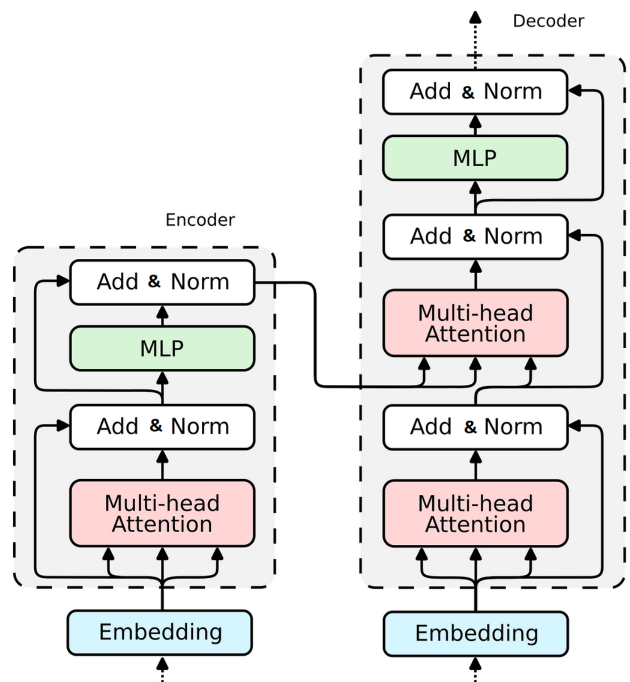

Figure 3. Transformer architecture [11].

After that, the transformer architecture is used to train a large language model or LLM (a kind of auto-regressive model that predicts the next token of a large collection of texts) and results in the first generation of generative pretrained transfomers (GPT1) [25]. GPT1 is trained on the BookCorpus [26] dataset which consists of around 7000 books and a total size of 4.5 GB (the training data was not filtered). The training of GPT1 is done in two steps: first, the LLM is trained on unlabeled text (the self-supervised stage) to learn the initial version of the neural network weights; then, the network is fined-tuned to a specific target task using supervised learning. Consequently, the pretrained weights are re-adapted to the new task (which is mainly instruction following). The new versions such as GPT2 and GPT3 have a larger size of weights and a variety of data sources with respect to the first

1. https://web.archive.org/web/20200907212153/http://image-net.org/about-stats.php

version. More specifically, GPT2 is trained on 40 G.B of human filtered text data and has a total of 1.5 billion parameters; while, GPT3 is trained on 450 G.B of human filtered data and has a total of 175 billions parameters. We notice that ChatGPT is an LLM that leverages either GPT3.5 or GPT4. It is trained in two stages: (1) an unsupervised learning for the auto-completion problem, (2) a combination of supervised learning and a reinforcement learning for the instruction following task.

3. Some A.I's achievements

In this section, we present various A.I systems that created breakthroughs and bypassed the human capabilities in a specific field, we will skim over many fields such as object recognition, speech recognition, reinforcement learning (including complex games), and protein folding.

3.1. Object recognition

Starting from 2015, deep neural networks including residual networks [19], efficient networks [20] (EfficientNet), EfficientNet-L2 [27] and other models bypassed the human performance in identifying the main objects and categories present in images. We notice that the human performance is equal to an error rate of 5.1% on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Figure 4 [28] shows the performance of a set of machine learning models in the ImageNet challenge. We notice

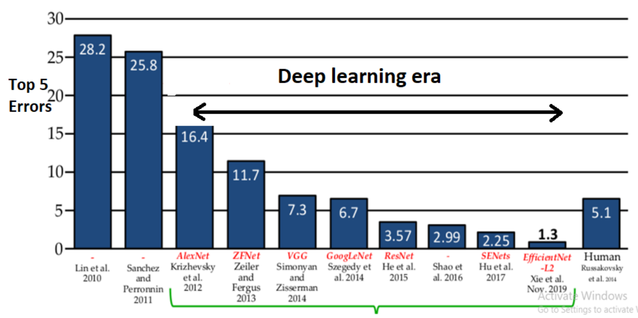

Figure 4. A.I models performance on ImageNet [28].

that the first breakthrough was ensured by residual networks of Microsoft. This architecture was able to reach this success level through the use of residual blocks (RB). RB is a composition of 02 convolutional layers (each one is a sequence of a dot product and a Relu activation function) in which we add a skip connection that connects the inputs of RB to the the Relu function of the second convolutional layer. The main motivation of RB is to allow the learning of identity functions that ensure the convergence of deeper CNN and consequently reaching a satisfactory training error.

3.2. Speech recognition

The automatic speech recognition (ASR) field has experienced many A.I models that include connectionist temporal classification (CTC), attention-based encoder decoder, recurrent neural networks (RNN), fully convolutional networks (FCN), transformers, and their combinations. The current models have an exceedingly increasing performance on well-known datasets such as the challenge of the Wall Street Journal dataset (WSJ) and the Librispeech dataset. We notice that WSJ is an english clean speech database having 80 hours, whereas, LibriSpeech is a dataset of reading speech from audio-books that contains 1000 hours of data. The following table shows a subset of the most effective models on ASR.

TABLE 1. PERFORMANCE OF SOME ASR MODELS.

Authors	Datasets	Word-error-rate Clean validation set	Word-error-rate Other set
RNN: [29]	LibriSpeech dataset	5.4%	15.6
	WSJ dataset	5.1%	8.4%
FCN: [30]	LibriSpeech dataset	3.08%	9.94 %
	WSJ dataset	3.5%	6.8 %

3.3. Reinforcement learning and games

The GO game (see Figure 5) is one of the most complex games in the history of humanity, its board contains 19 lines and 19 columns, and therefore there are 19*19 possible configurations to explore. In addition, the average branching factor (the number of moves) is 150. It is worth noting that the number of legal configurations of Go is close to 2.1×10^{170}, and this amount is higher than the number of atoms in the observable universe [31].

Figure 5. The board of Go.

In 2016, an A.I model based on deep reinforcement learning (AlphaGo) [32] of DeepMind company was trained on the Go game. The AlphaGo model leveraged deep convolutional networks and beat the world champion Lee Sedol with a score of 4-1.

3.4. Protein folding

In 2021, The deep neural network Alpha-fold of DeepMind succeeded in resolving the problem of protein folding (PF). PF is the shape prediction of protein from the amino acid sequence. Alpha-fold [33] wined the standard competition of the issue known as Critical Assessment of Protein Structure Prediction (CASP 14th edition). Alpha-fold scored 244.0 compared to 90.8 which is the score gained by the next best group. The predictions made by AlphaFold had a median success rate of 92.4%.

4. A.I challenges and threats

Both narrow A.I and AGI are facing many challenges that are far from being positively handled at the current state. Likewise, some potential A.I threats are also pending and must be addressed by the both the research community and decision makers. As regards the challenges, we give an overview about the main obstacles that we face today:

- **A.I bias and ethical issues:** the training data used in large language models and GPT like models can be unbalanced and ethically deviated (since it reflects the convictions and hidden thoughts of materials' owners) and this lead to biases in the learned knowledge. For instance, A.I may generate negative content about specific ethnic groups or specific religions and this reinforces divisions in the same society.
- **Robust A.I:** the ability of learning reliable models despite the lack of labeled data, presence of noise, and contradictory data present in the web is a difficult task; in addition, the adversarial attacks may change the training examples in order to mislead the future decisions of A.I systems. In this perspective, the use of multimodal data is a way of producing reliable predictions and compensating the lack of labeled data, especially in medical fields. In the medical domain, the integration of multimodal inputs not only enhances the accuracy of the diagnosis but also provides valuable explanations about the automated decisions and the fore-castings. Moreover, theoretical frameworks such as probably-approximately-correct learning (PAC-learning) [13] constitute a solid way for dealing with the issues of robust I.A, but the theoretical conditions of the PAC framework are hard to meet in practical cases.
- **Auto machine learning (AutoML):** the ability of creating machines that not only learn good models for specific tasks, but also automatically fine-tune the hyperparameters including the activation functions, kernels' configurations, rearrangement of the cost function landscape. Moreover, AutoML involves the capability to perform transfer learning from previously learned problems to a wide range of new tasks. This human inspired ability allows for saving the time and resources of retraining models from scratch and therefore these past skills can be re-adapted to other problems and fields.
- **Limited resources and energy:** deep learning models and especially large language models (such as GPTs) are highly greedy in terms of computing power and time of training; in this regard, the acquirement of resilient data centers with permanent energy of functioning and cooling is a prime necessity. This issue is exacerbated with the expensive and fluctuating prices of hydrocarbon-based energy; moreover, the green energy is still insufficient and unavailable in some periods due to environmental and atmospheric constraints.

Concerning the A.I threats, it is well known that the possible dangers may arise from the misuse of of A.I technologies. In what follows, we mainly cover the prominent dangers that can affect the humanity and their life style.

- **A.I misuse and fake information:** One of the main misuse of A.I systems is the utilization of collected data about users (of mobile applications, web applications, and social networks) and the deduced knowledge to harm specific ethnic groups or minorities. In the same line of thought, some generative models may create fake data (images, audios, or videos) which are almost indistinguishable from real data, and stick them to other entities in order to defame or harm them. these fake data can also be used to falsify evidences and deviate the investigation paths in case of crimes or dereliction.
- **Job loss:** Traditionally, robotics and narrow I.A have tendencies to replace human workers by automated machines in routine tasks (e.g., manufacturing, mail filtering and ranking). But the category of target tasks is increasingly expanding, and it involves tasks such as data analysis or document review, newspaper report writing, novel writing, research paper writing, and medical imaging analysis. Consequently, job positions in domains like finance, lawyering, and journalism are at high risks. In the same regard, we observe that workers who have low-quality (and sometimes mid-quality skills) in some fields such programming are at risk, and they can be laid off by their companies and replaced by automated tools. This can lead to a growing gap between high-skilled and low-skilled workers and create an economical issue since the unemployment rate increases.
- **A.I and privacy concerns:** Nowadays, many A.I systems collect private and other sensitive data about users and institutions without their agreement. This type of data can be sold to third parties or can be used for training A.I models that can harm people and organizations in the future. Another concern is that even authorized institutions (such as hospitals) can use private data and learn vulnerable models that can be attacked by malicious entities. In this regard, model inversion attacks and membership inference attacks can be applied on vulnerable A.I models and infer sensitive information. To tackle these risks, researchers combine differential privacy [34], secure multiparty computation [35], federated learning [36], and homomorphic encryption [37] to alleviate the menaces.

5. Conclusion

In this paper, we have presented the definitions and the perfection levels of A.I. We have also reviewed the various stages of the A.I progress (From the Turing Test up to the actual GPT's). Furthermore, we have shown the impressive results of some existing A.I models and have given an overview over the potential dangers of the future machine intelligence. The next questions must focus on the possible improvement and adaptation of future AGI Systems in our daily life and environment, especially in healthcare, smart cities (e.g., transport, smart buildings, energy generation, and government), agronomy and industry.

References

[1] L.S.Gottfredson. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Intelligence 24, no. 1. 1997. 13-23.

[2] R. Kurzweil The singularity is near. In Ethics and emerging technologies London: Palgrave Macmillan UK. 2005. 393-406.

[3] B.Goertzel. Artificial general intelligence: concept, state of the art, and future prospects. Journal of Artificial General Intelligence, vol. 5, no 1, 2014. p. 1

[4] D.E. Rumelhart, G.E. Hinton and R.J. Williams. Learning internal representations by error propagation. Chapter 8 in Parallel Distributed Processing: Foundations. Vol. 1, MIT Press, Cambridge, MA,1986. 318–362.

[5] C. Corinna, and V.Vapnik. Support-vector networks. Machine learning 20 .1995. 273-297.

[6] L.Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees, Wadsworth, Belmont, CA. Republished by CRC Press. 1984.

[7] Y.LeCun, L. Bottou, Y.Bengio, and P.Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86, no. 11. 1998. 2278-2324.

[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems 25. 2012.

[9] H.Sepp, and J. Schmidhuber. ”Long short-term memory.” Neural computation 9, no. 8. 1997. 1735-1780.

[10] K.Cho, B.Van Merriënboer, C. Gulcehre, D.Bahdanau, ,F. Bougares, H.Schwenk, and. Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.2014.

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and I.Polosukhin. Attention is all you need. Advances in neural information processing systems, 30. 2017.

[12] D.Wlodzislaw, R.J. Oentaryo, and M.Pasquier. ”Cognitive architectures: Where do we go from here?.” Agi. Vol. 171. No. 1. 2008.

[13] L.Valiant. Probably approximately correct: natureõs algorithms for learning and prospering in a complex world. Basic Books (AZ), 2013.

[14] Y.Freund, R.E.Schapire: A decision-theoretic generalization of on-line learning and an ap- plication to boosting. Journal of Computer and System Sciences 55(1), 1997. 119–139.

[15] L.Breiman. Random forests. Machine learning 45. 2001. 5-32.

[16] L. E.Baum; T.Petrie ”Statistical Inference for Probabilistic Functions of Finite State Markov Chains”. The Annals of Mathematical Statistics. 37 (6). 1966. 1554–1563. doi:10.1214/aoms/1177699147.

[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

[18] C.Szegedy, W.Liu, Y.Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition.2015. (pp. 1-9).

[19] K.He, X. Zhang, S.REN, J.Sun. Deep residual learning for image recognition. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

[20] M.TAN, Q.LE. Efficientnet: Rethinking model scaling for convolutional neural networks. In : International conference on machine learning. PMLR, 2019. p. 6105-6114.

[21] O.Ronneberger, P.Fischer, and I.Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[22] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017 Apr 27;40(4):834-48.

[23] D.Bahdanau, K.Cho, and Y.Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.2014.

[24] J.Cheng, L.Dong, and M.Lapata. ”Long short-term memory-networks for machine reading.” arXiv preprint arXiv:1601.06733.2016.

[25] A.Radford, K.Narasimhan, T.Salimans, I.Sutskever. Improving language understanding by generative pre-training. 2018.

[26] Y.Zhu , R.Kiros, R.Zemel, R.Salakhutdinov, R. Urtasun, A.Torralba, S.Fidler, S. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international conference on computer vision. 2015. (pp. 19-27).

[27] Q.XIE, M.T. LUONG, E.HOVY, et al. Self-training with noisy student improves imagenet classification. In : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 10687-10698.

[28] L.ALZUBAIDI, J.ZHANG, A.J.HUMAIDI, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, vol. 8. 2021.p. 1-74.

[29] T.Hori, J.Cho, S.Watanabe. End-to-end speech recognition with word-based RNN language models. In 2018 IEEE Spoken Language Technology Workshop (SLT). 2018. (pp. 389-396).

[30] N.Zeghidour, Q.Xu, V.Liptchinsky, N.Usunier, G.Synnaeve, R.Collobert. Fully convolutional speech recognition. arXiv preprint arXiv:1812.06864. 2018.

[31] ”A Brief History of Go — British Go Association”. www.britgo.org. Retrieved 2023-12-16.

[32] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. nature 550, 7676. 2017. 354–359.

[33] J.Jumper, R.Evans, A. Pritzel, T.Green, M.Figurnov, O.Ronneberger, K.Tunyasuvunakool et al. ”Applying and improving AlphaFold at CASP14.” Proteins: Structure, Function, and Bioinformatics 89, no. 12 .2021. 1711-1721.

[34] C.Dwork. ”Differential privacy.” In International colloquium on automata, languages, and programming, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.pp. 1-12.

[35] A.C. Yao. Protocols for secure computations. In Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, 1982.

[36] B. Keith, V. Ivanov, B. Kreuter, A. Marcedone, H. Brendan McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth. ”Practical secure aggregation for privacy-preserving machine learning.” In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,2017. pp. 1175-1191.

[37] R.L. Rivest, L.Adleman, and M.L. Dertouzos. On data banks and privacy homomorphisms. Foundations of Secure Computation 4, 11. 1978, 169–180.