Deep learning advancements: closing the gap

A. Stipić[*], T. Bronzin[*], B. Prole[*] and K. Pap[**]

[*]CITUS, Zagreb, Croatia

[**]University of Zagreb/Faculty of Graphic Arts, Zagreb, Croatia

arian@citus.hr, tbronzin@citus.hr, brigita@citus.hr, kpap@grf.hr

***Abstract* - This article explains how recent development in the field of Artificial Intelligence (AI) makes gap between human and machine smaller than ever before, by explaining and comparing traditional approach user in development of AI systems with new approach that has been used by AI system AlphaZero, developed by DeepMind. Traditionally AI systems have been tested in chess and the same has been done to demonstrate the power of AlphaZero. But, instead of playing against human, it played against the best (at the time) chess program Stockfish. While chess programs (before AlphaZero), were using powerful hardware and embedded built-in formal knowledge about the game, AlphaZero is using completely new approach, running on standard hardware and using deep learning. It learned about the game by playing a large number of games with itself, learning in the process. Article will also explain what is so revolutionary in AlphaZero approach to AI and how this new approach can be used in different areas of processing visual information, bio-medicine, autonomous driving, robotics and AI generated images/videos of humans.**

Keywords – artificial intelligence; deep learning; AlphaZero; bio medicine; autonomous driving; GAN; generated image; video synthesis;

I. INTRODUCTION

The idea of artificial intelligence as a concept that would be able to somehow emulate the functionalities of human brain has been fascinating mankind for a long time. We have been (and, for the most part, still are) wondering what exactly the difference between the inner workings of a brain and a very advanced machine is. These days, or at least from 1950, the machine really means a computer of some kind.

For a long time, the question did not have much practical value because computers could not really compete with humans in almost anything, but that was only a consequence of available "firepower" or the sheer computing power. However, things have gradually started to change, hardware was more and more powerful and mathematical formalism used to approach the topics of artificial intelligence has also advanced to the point where question began to make sense.

II. ARTIFICIAL INTELLIGENCE (AI)

A. Setting the stage

At the same time people realized that there are also many other important questions that were (and many still are) waiting to be answered. As the computer's capabilities increased, the number of different important areas have also increased and natural language processing, handwriting recognition, speech recognition, image recognition, games playing, advanced robotics, evolutionary systems, autonomous driving and many other problems were elevated to the level where we started looking for a well-defined, logical and mathematical answers. And so, the discipline of AI (artificial intelligence) has been officially established.

Despite big promises, things were not always smooth and the society (even very qualified parts of it) did not always recognize that this is yet another new technology and it should, therefore, follow the similar development and adoption path as other important technologies do. This is nicely explained in the following Gartner's image from 2017, Figure 1:

Figure 1: Gartner, 2017

AI did not exactly follow this pattern. Instead, it moved along this one, Figure 2:

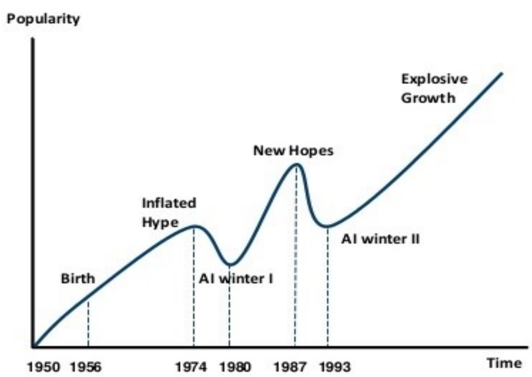

Figure 2: Team Finland Future Watch Report, 2017

If you take a good look at the above image, you will be able to see that it looks like there was not one but two starts (inflexion points) instead and, very important, that the rapidly rising curve is much steeper than Gartner would predict. But otherwise, the pattern is there, and we should be able to analyze it in a rational manner.

In a sense, for a long time (more than 60 years, actually) AI has been a victim of our reduced capabilities and it simply means that we did not have enough computing power to investigate the theoretical possibilities and, therefore, we were very far from being able to build important commercial level applications that would benefit from AI promises. To be completely fair about this, there were quite a few important and even impressive steps, like IBM's DeepBlue that won the first chess match against Gary Kasparov in 1997. That marked the end of an era of human domination in probably the most beautiful logical game that human race ever invented.

But these successes were still very limited and were certainly still far from becoming a part of our standard productivity applications. The hardware was still too expensive, the solutions were too specialized for a certain task and something was clearly still missing.

In this period many reputable teams were working on natural language processing and speech recognition, and some have also started to develop image recognition techniques because these were the areas that seemed to promise the most at the time. Some important examples that even entered the commercial stage were Roomba (the first autonomous cleaner that is widely available today) in 2002 and Apple's Siri (speech recognition intelligent assistant) in 2011. The time difference between these two was very typical for AI struggles in those days.

At the time, IBM was still a very strong player in the field of AI (they would surely argue that they still are, but not everyone agrees with that) and they proved it by building yet another specialized machine that was able to beat men in the jeopardy game in 2011 and that was an important step towards fulfilling Adam Turing's three criteria for comparing AI to human brain.

B. Deep learning revolution

Waiting for more capable hardware lasted for another 15 years and new ideas have been boiling. If you look back at those years at the beginning of 2010's, two very important things were emerging and the other one really played a key role for rapid development of AI and deep learning in particular. That other thing was, of course, Big Data based on Hadoop and other new related technologies. From the very beginning Big Data (compared to AI) had a clear commercial value and very soon it became the main topics in analytics everywhere. It brought two important breakthroughs to the table: the ability to process very large (petabytes and even more) data sets and, also extremely important, the ability to work with and analyze unstructured data like text, speech and images.

The important new page for AI has been turned when Geoffrey Hinton (and his colleagues David E. Rumelhart and Ronald J. Williams) have promoted an old idea of multi-layer neural networks, which opened the field that is today known as deep learning [1]. But the real breakthrough came when Alex Krizhevsky designed the innovative solution for the Imagenet challenge in 2012 that helped to revolutionize the field of computer vision [2].

Now, the stage was finally ready for the most important breakthrough in AI: deep neural networks and deep learning concept. In other words, the idea that a machine can be programmed to learn from its own experience and not something that has been somehow preloaded, such an idea has become the most important path of development to come.

Deep learning has been developed as the extension of ordinary neural networks and the idea of another concept called "unsupervised learning". Basically, the idea is to supply a minimum number of parameters, feed a significant amount of representative data and then let the computer go through a process called "training" to build a model that can then later be used with other production data to predict some aspects of future behavior of the real-life system. The beauty of this concept is that, while training and building a model usually still require very significant computing power, using a model (once it has been trained) can be very effectively done on ordinary, commodity hardware. And that, of course, makes it extremely interesting for a variety of commercial applications.

Many different algorithms are used in deep learning coupled with some of the most advanced mathematical concepts such as tensor analysis, and they are not trivial. Tensors happen to be a good foundation for multi-layered neural networks. That is because, at the very basic level, neural networks consist of these layers:

- Input layer
- A series of hidden layers, and
- Output layer

This is represented on the following image, Figure 3:

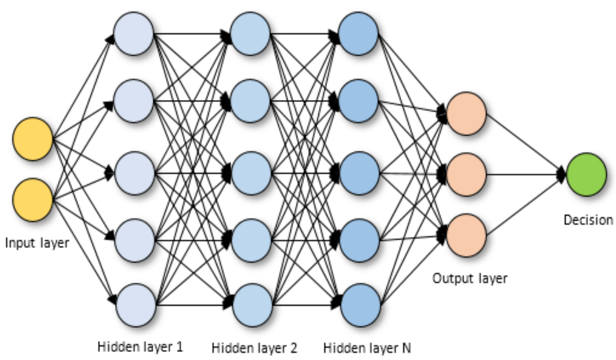

Figure 3: Multi-layer processing in neural networks

The number of hidden layers can be very high, depending on the number of different characteristics and the overall complexity of the problem (process) that we want to solve.

Once you have trained your model, it can be put to work and, in principle, the process starts with the most basic elements, like identification of individual pixels, colors and shapes. After that, each of the subsequent layers processes higher-level (more complex) elements until enough information has been collected to calculate the output probabilities in the output layer and make the final decision.

The following image represents this process in some detail, Figure 4, [3]:

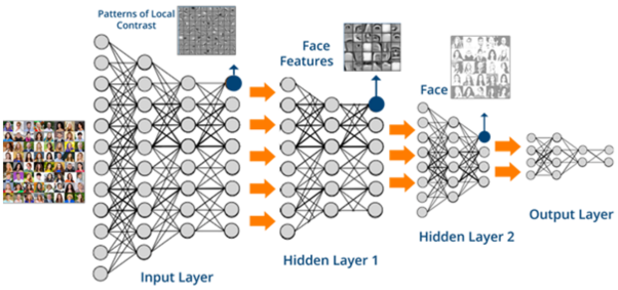

Figure 4: Multi-layer image processing

Deep learning has quickly gained popularity for two important reasons:

- Models were able to work with huge amounts of data and still keep high performance levels
- People discovered that the very same (or at least very similar) models could be applied to a wide variety of entirely different problems and still get very good results

Here we can immediately see the link to Big Data developments, something that will quickly become one of the cornerstones of deep learning developments. It is not difficult to understand the importance of Big Data for AI because:

- Most of the data that AI needs to analyze is usually not structured but instead, AI works with pictures, sounds, text and similar data types
- The amount of data available for describing practically important problems tends to be extremely high, not only in training phase, but also in the production phase

Some examples of such problems are complex games, autonomous driving, mixed and simulated reality, processing of LHC (Large Hadron Collider) images, chemical and bio-medical problems, evolutionary based systems and many others.

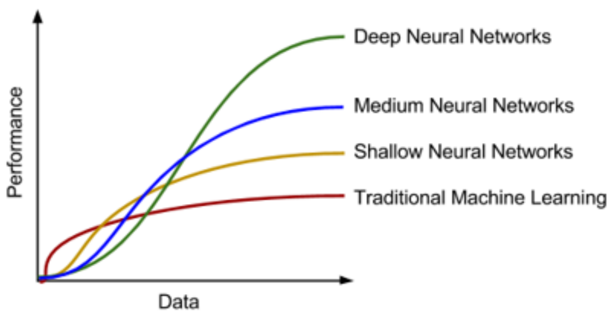

Figure 5: Performance vs Data Volume

There is another very important characteristic to be noticed here. There are many different types of neural networks and not all of them are equally friendly to large amounts of data. Common sense might lead you to the conclusion that the best performance should be expected from the simplest neural networks, or at least from those with least number of hidden levels. Interestingly enough, this is not the case. In fact, it is exactly the opposite, as described on the previous image, Figure 5 [4].

This unexpected result shows that more complex, deep neural networks, actually favor huge amounts of data. The more data you supply, the better performance you get! Obviously, this type of behavior immediately makes the deep neural networks everybody's favorite for solving complex problems.

C. *AlphaZero and Leela*

It is almost impossible that you have not heard of AlphaZero and its chess match against Stockfish in 2018. Stockfish was practically undisputed world chess champion (among computers; humans cannot compete with them for quite some time now) and it has been built using totally different approach, not based on AI.

Chess world has a concept called ELO rating, an official mathematical system that allows objective classification for all players, including computers! Stockfish 8 was rated at 3300, far above human world champion. And what about AlphaZero? Well, based on this extremely narrow sample of games, it has been estimated to be at least 4100 points, but more likely to be much higher than that.

Just consider this fact for a second: the difference between best human and best machine is roughly 400 points and humans cannot compete with computers. The difference between AlphaZero and one of the best machines seems to be at least 800 points, possibly even several times more.

If you are interested to learn more about AlphaZero, there are many articles about that match available on the internet and a good place to start with is here [5].

DeepMind released the paper soon to be followed by with initial 10 games, very detailed statistics about all played games and revealed the underlying technology used in AlphaZero [6].

Although there were only 10 games originally released, it instantly became clear that something really extraordinary has happened. Although AlphaZero was superior in that match, the main point was not the result itself. Instead, it was how some of the games were played and the general consensus around the world was that it played not like a computer, but like a super-human instead. That really meant that the style was like a human but a very, very advanced one. How advanced, exactly? Well, final score in the 100 games match was: 28 wins for AlphaZero, 0 losses and 72 draws and the rough ELO estimate was set to 4100 points or higher.

This level of supremacy made some people very suspicious and some of them still are! Although the very same program previously defeated human champions in Go and Shogi (two most complex games invented by humans), this was something very different because the defeated party was another computer program. And this other program was so superior to humans that it has been considered as practically invincible.

So, what is it so outstanding about the chess match between two machines?

First of all, there is a difference in how the two programs work. Stockfish 8 (like practically all other programs of its kind) uses a lot of built-in formal knowledge (opening theory, endgame tables, specific chess related knowledge) and combines all that together while playing a game.

AlphaZero does something completely different. The only thing that it knows about the game (Go, Shogi or Chess) is the complete set of game rules. Nothing else whatsoever. No theory, no principles, no endgames… nothing at all! The only thing that it had to do is to somehow "learn" how to play good chess and it does so by first playing a large number of games with itself. This was the training process and AlphaZero took only 9 hours to train itself from zero to demonstrated strength!

So, what is the big deal about that?

To start with, that it is not about the processing speed. In fact, AlphaZero is analyzing roughly 80 thousand positions per second, while Stockfish 8 was analyzing much higher number, over 70 million. Normally, this should give Stockfish 8 significant advantage, but the reality was completely the opposite. So, if it is not about the processing power, it must be about the principles of processing. The two programs are clearly not "thinking" the same way. In other words, they do not share the same level (even not the same type!) of intelligence.

Today we have more than 100 published games and it is very clear that initial conclusions were right on the target.

But, the real test for AlphaZero technology came from different side after the results have been published. The key question was: can the same principles be successfully applied using commodity hardware instead of super-expensive Google TPU (tensor processing unit) processors?

It did not take long and, based on the information published by DeepMind, Gary Linscott have created special version of AlphaZero for ordinary PC's. The new program was named Leela Chess Zero (or simply: Leela) and it started with zero games played. He did not have powerful machines to do the training so instead, he asked people around the world to play with Leela and create the necessary large number of training games. In December 2018, the AlphaZero team published a new paper in Science magazine revealing previously undisclosed details of the architecture and training parameters used for AlphaZero. These changes were soon incorporated into Leela Chess Zero and increased both its strength and training efficiency.

In February 2019, Leela scored first major tournament win in the second TCEC cup, an unofficial world championship for chess computers. She defeated the reigning champion Houdini in the final match of 8 games. Even more important, Leela did not lose a game throughout the entire tournament!

This really was a definitive empirical proof that deep learning methodology introduced by DeepMind in their AlphaZero was one of the most important achievements in the history of AI and history of deep learning in particular. Important point here is that very similar implementations can be used for entirely different types of problems that have nothing to do with chess at all.

It brings us to summarize our conclusions about the state of deep neural networks today, which can be simplified to these three key points:

1. Hardware has become powerful enough and even the commodity hardware can support quite advanced AI scenarios
2. Data is being generated very fast, in many different forms and places and in huge volumes; Deep neural networks love this and are able to keep-up with this very successfully
3. Algorithms and tools are being developed using very advanced mathematical constructs that dramatically improve the power of AI

Old questions about the AI and its potential have risen again, this time often in a much more serious formulation. While it is true that machines still don't have human's general cognitive capabilities, it is not at all clear where the limits might really be. In fact, many reputable people seriously think that. If not handled properly, one day AI might become the greatest threat to mankind in general. On the other hand, many believe that it will help us solve many complex problems that our brains cannot solve today.

The gap is closing…

D. What is next

There are several areas where AI can be used. Potentially the most interesting applications are in the field of real-time, or near real-time image or video processing and intelligent decision making. This paper has been focused on the state-of-the-art advancements in the area of computer-generated images and videos, that can be used in exciting new disciplines like autonomous driving, various bio-medical applications of deep learning, processing extremely large volume of images generated in physical experiments in LHC (Large Hadron Collider) or creating humanoid avatars and robots.

Computer vision is one of the most advanced area of usage for deep learning, especially in the 2017-2018, when number of scientist, including research team led by DeepMind experts, have written/published several scientific papers about application of deep learning in the area of bio-medicine.

One of the papers that has been written by DeepMind researchers (and their collaborators) is about their AlphaFold AI powered system that has been used to predict 3D protein structure, based on the knowledge from previous researches and huge quantity of genomic data, more accurate than any that have come before [7]. The importance of determining the right 3D sequence of how the proteins are folded is important because it can help in diagnosis and curing diseases that has been suspected to be related to misfolded proteins like Alzheimer's, Parkinson's, Huntington's and cystic fibrosis. The other area of usage is design of new proteins that can be used, for example, in the plastic waste management, where proteins can help develop biodegradable enzymes that helps to fight pollutants like oil and plastics. Artificial intelligence, namely deep learning makes the process of determining the 3D structure of proteins faster and chipper, as previous methods were based on trial and errors using different experimental technics. In their paper [7], DeepMind team explained how they prepared their submissions to the global competition called the Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP), marked as A7D. For purpose of producing submissions to the CASP, three variants of an automatic free -modelling structure prediction system were used that has been relying on scores computed with deep neural networks [7]. For the purpose of this paper, most important facts are that they *"... didn't use any templates or server predictions and no manual intervention was made except for domain segmentation of T0999 and final decoy ranking in a handful of cases"* [7].

The second paper published by team of scientists led by DeepMind researchers explains applications of deep learning in field of diagnostic imaging [8]. DeepMind researchers focused on achieving performance in three-dimensional diagnostic scans, namely three-dimensional optical coherence tomography (OCT) scans from eye hospital patients. In their research, DeepMind team used small number of manually selected segmented tissues for training purposes that resulted in fast and very reliable diagnosis very close to what the human experts are capable to provide. Furthermore, they removed dependency on the type of devices that has been used for image tissue segmentations that, in the past, required modification of the model used for tissue segmentation. This alone can help human experts today and speed-up diagnostic process, but future use of this or similar methodology and technology promises to solve the "second step": to achieve expert performance of the proposed AI solution in providing referrals decisions.

Another big area of deep learning application is in the field of autonomous driving, either cars or drones or commercial jets.

In the past two years number of announcements about self-driving cars, beside Tesla, has been made by Swedish car maker Volvo that hopes to introduce a fully autonomous commercial vehicle in 2021 [9] and BMW & Daimler [10]. Other leading car-makers and IT companies in self-driving car industry have recognized the importance of self-driving cars for the future of transport [11].

Self-driving cars has been in the focus of the researchers for the long time, especially in last few years when so called "human-centric" approach has been emphasized as the promise of fully autonomous driving has not been reached yet. Another reason for taking "human-centric" approach is question of ethics and broad discussion in the society related to the first (human) fatal incidents by Tesla [12] and Uber [13] self-driving cars. In contrast to the human-centric autonomous driving, traditional approach would be to completely ignore human as a driver and concentrate on solving technical and planning problems.

Human centric approach to the autonomous driving represents so called shared-autonomy, in contrast to ideal fully autonomous vehicles. Several talks and papers have been published about this topic, among them two that are talking about symbiotic autonomous vehicles (SAV) *"... which can work with human users in a safe way, leveraging the human/machine perception, decision making and action when needed.*" [14] and Human-Centered Autonomous Vehicle Systems (HCAVS) [15] where humans and machines interact during the driving.

Authors define HCAVS with seven principles of shared autonomy: Shared Autonomy, Learn from Data, Human Sensing, Shared Perception-Control, Deep Personalization, Imperfect by Design, System-Level Experience [15]. Although the responsibility to control the vehicle stays on the human driver, about 30% (Tesla Autopilot is a good example here) of automated tasks and actions can be done by AI. HCAVS uses AI, namely supervised learning to achieve further improving to level 50% of automated response, going to semi-supervised and unsupervised learning formulations [15]. Data from human driver sensing, like gaze, hand position and pose are part of the decision process too. HACVS can enable better results in today-ready semi-autonomous vehicles, while we wait for fully autonomous one.

Another interesting area of using AI is creating completely new video based on existing video, so called video-to-video synthesis problem "*... as a distribution matching problem, where the goal is to train a model such that the conditional distribution of the synthesized videos given input videos resembles that of real videos.*" [16]. Video-to-video synthesis is used in several areas, including: generation of a high-quality video from an input segmentation map, edge-two-face video generation, pose-two-body video generation and frame prediction. All of them are built on Generative Adversarial Networks (GAN). GAN are two separate neural networks, one is generator, used to produce realistic synthetic data, and the other one is discriminator, used to analyze videos syntheses by generator. If discriminator is fooled by generator (can't differentiate synthetic generated frames from the real ones), then the process is successful. New video-to-video synthesis method proposed by [17] use conditional video generation with GAN's [17] that allows flexible control over the output of the model, that "*… synthesizes photorealistic videos conditioning on manipulable semantic representations, such as segmentation masks, sketches, and poses."* [16].

Research results that have been gathered from video-2-video syntheses can be used in generation of new virtual human characters. Recently, very popular application is to create virtual humans on television as TV anchors. Those virtual humans represent completely new persons, use gestures and creates stories.

Finally, humans like to communicate with other humans. Because of that, some of the most active research of AI applications are in robotics. In spite the fact that event the self-driving cars are robots, and that robots are used in industry/manufacturing, the most advanced uses of AI, at the moment, is for development of human-like robots.

III. Conlusion

The beauty of AI specifically deep learning is that it is applicable to the broad range of problems. If you implement enough number of layers and methods, it's possible to solve huge number of different complex problems. In contrast to the traditional machine learning, deep learning is performing better with more data available for training the model/learning. On the other hand, deep learning performs well even in cases when unsupervised learning is used for training the model.

Also, deep learning exhibits new capability that is close what humans call “intuition” that helps in solving problems with better efficiency. In spite of number of breakthroughs and advancement in AI when we are speaking about deep learning, generally success of AI application in specific field depends on the fact how well the problem that we want to be solved is defined and if we have chosen the right problem to solve.

Yet, if we have well defined the problem and we had chosen the right tools for solving the problem, then we need to set the right goals that we want to achieve. For example, when researchers started to work on developing autonomous vehicles, most of them have been focused on full self-driving cars. In the meantime, another approach has been receiving more and more support – human centric approach that is focused on understanding how human thinks, behaves and reason. That leads to the recent AI development that has been focused on improvement of human-to-machine collaboration and communication in symbiotic way.

In the field of bio-medicine, problems are better defined, goals are set, and results are better, but they still need to pass clinical testing. On the other hand, when human-centric approach is applied, meaning AI collaborates with humans, the best results are achieved. This conclusion has ground in example of using AI to produce device-independent scanning diagnostic of tissue so human medical experts can establish referral to diagnosis, with important additional tools: ethics and empathy.

Using AI to create virtual humans that are not avatars, create fake videos, by using video-to-video synthesis, or even make robots more human makes “real humans” anxious and afraid what the future AI development will bring. Because of that, it’s important to promote further discussion about ethics, regulations and law’s in all areas where AI is about to be used.

References

[1] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors", Nature. 323 (6088): pp. 533–536. doi:10.1038/323533a0. ISSN 1476-4687, 1986

[2] A. Krizhevsky, I. Sutskever and Ilya; G. E. Hinton, "ImageNet classification with deep convolutional neural networks". Curran Associates Inc., pp. 1097–1105, 2012

[3] Graph initialy published at: https://puentesdigitales.com/2017/11/15/guia-rapida-de-inteligencia-artificial-asi-funciona-el-deep-learning/, February 2019

[4] Graph initialy published at: https://blog.easysol.net/building-ai-applications/, spoted on February 2019

[5] Article initialy published at: https://www.linkedin.com/pulse/alphazero-ai-huge-consequences-arian-stipi%C4%87/, spoted on February 2019

[6] D. Silver, T. Hubert, J. Schrittwieser; I. Antonoglou, M. Lai, et al., "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm", published at: https://arxiv.org/abs/1712.01815, 2017

[7] R.Evans, J.Jumper, J.Kirkpatrick, L.Sifre, T.F.G.Green, et al., “De novo structure prediction with deep-learning based scoring”, In Thirteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstracts) 1-4 December 2018. Retrieved from here: https://deepmind.com/documents/262/A7D_AlphaFold.pdf, unpublished

[8] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, et al., “Clinically applicable deep learning for diagnosis and referral in retinal disease”, Nature Medicine, volume 24, pp.1342–1350, https://doi.org/10.1038/s41591-018-0107-6, 2018

[9] Initially published at: https://scipol.org/content/volvo-takes-%E2%80%98human-centric%E2%80%99-approach-driverless-cars, spoted on February 2019

[10] Initially published at: https://www.eenewsautomotive.com/news/bmw-daimler-join-forces-automated-driving-technologies, spoted on February 2019

[11] Initially published at: https://www.zdnet.com/article/dossier-the-leaders-in-self-driving-cars/, spoted on February 2019

[12] Initially published at: https://www.2025ad.com/latest/tesla-fatal-crash-analysis/, spoted on February 2019

[13] Initially published at: https://www.2025ad.com/latest/uber-driverless-car-fatal-crash-analysis/, spoted on February 2019

[14] P. Boyraz, “Is it Possible to Rethink Autonomous Vehicles with Human-centric Design Principles? Taming the AI-powered Robotics with Human-Factors and Safety”, Conference: IEEE SMC- Symbiotic Autonomous Systems Workshop, Presentation, http://bit.ly/mipro2019_16, 2018

[15] L. Friedman, " Human-Centered Autonomous Vehicle Systems: Principles of Effective Shared Autonomy ", published at: CoRR, volume 1810.01835 https://arxiv.org/abs/1810.01835, 2019

[16] T. C. Wang, M. Y. Liu, J. Y. Zhu, G. Liu, A.Tao, et al., “Video-to-Video Synthesis”, Advances in Neural Information Processing Systems (NeurIPS), https://arxiv.org/pdf/1808.06601.pdf, 2018

[17] T. Karras, S. Laine, T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks”, Cornell University, arXiv:1812.04948v2 [cs.NE], https://arxiv.org/abs/1812.04948, 2018