

# A Review of Recent Trends and Industry Prospects for Artificial Intelligence Technologies

Guangjun Li  
University of Technology Sydney  
Sydney, Australia  
liguangjun1117@163.com

Xianzhi Wang  
University of Technology Sydney  
Sydney, Australia  
xianzhi.wang@uts.edu.au

Minxi Li  
L.F.-Sun Pty Ltd  
Melbourne, Australia  
minxilee1@gmail.com

**Abstract**—Recent years have witnessed various breakthroughs and advancements in research and application of Artificial Intelligence (AI) technologies. Notably, AI is leading the 4th global industrial revolution (Industry 4.0). In fact, deep learning rose to become one of the most highly researched aspects in various industrial sectors worldwide, making it especially important to timely review the recent progress and industry prospects of AI. In this paper, we briefly review the recent advances in AI technologies, discuss the various advantages and challenges they bring, and finally, offer a few points of thought regarding the future developments in the industry.

**Index Terms**—AI technologies, Industry 4.0, Deep Learning

## I. INTRODUCTION

The AI industry is projected to bear the Market size value of USD 93.53 billion in 2021, and this number is expected to continue rising in the coming year<sup>1</sup>. On the one hand, the industry keeps expanding the boundaries of deep learning to solve problems and gain cognitive enhancement of applications. While deep learning remains the leading route followed by artificial intelligence (AI) technologies, it faces significant challenges that limit its application in refined segmented scenarios. For example, current deep learning techniques heavily rely on the availability of sufficient labeled data [1]; this contradicts to data sizes of practical applications—a typical practical application requires over 100,000 labeled samples for training. For example, ImageNet dataset [2] contains more than 14 million images, and Open Catalyst contains more than 1.3 million intermolecular interactions of compounds.

On the other hand, current AI research has gone beyond purely supervised learning [3]. New trends of research focus on improving the quality and scale of data, overcoming the limitations of deep learning via transferring training results from other domains [4], generating or enhancing data autonomously [5], leveraging knowledge graph and common-sense knowledge [6], and taking advantage of multi-source data [7]. Meanwhile, the industry has witnessed notable advances and exhibited diverse development routes, represented by vertical integration with domains, platform-based development, and multiple forms of ecosystems. It continuously calls for research and practical innovations sustained by growing computing capabilities.

<sup>1</sup><https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market>

This paper is motivated by the fruitful recent advances in AI research, industry, and related technologies, as well as the lack of a comprehensive review or commentary that bridges AI research and industry in literature. We aim to present briefly review the current state of development of the AI industry, with a focus on the new trends, key technical components, pain-points, future directions, to cover this gap. The remainder of this paper is organized as follows: we briefly overview trending directions in current AI research and development (Section II); then, we offer a few outlooks into the AI industry (Section III); finally, we give some concluding remarks (Section IV).

## II. TRENDS IN ARTIFICIAL INTELLIGENCE

### A. Integration of Multi-perspective Innovations

The combination of deep learning and multiple learning fashions is increasing explored. Small-sample-size training [8], weak human intervention, and multimodal learning [9] have become desirable properties in research and development. Small-sample-size training (or few-shot learning [10]) aims to reduce the dependence on data volume. Few-shot learning uses a small amount of data to train new domains by reusing the knowledge structure of other domains; it has been preliminarily applied in practice, e.g., Nvidia's Few-shot learning-based video transformation (Few-shot vid2vid [11]) framework can synthesize videos of a target or scene that has not appeared with only a small number of sample images of the target or scene. Weak human intervention is primarily driven by the high cost of manual labeling in preparation of sufficient training data; and self-supervised learning [12] and reinforcement learning [13] help reduce human intervention in deep learning, counter-striking the high cost of data labeling in mainstream supervised learning methods in the AI industry. Take machine translation for example, the market-average price of human translation per word is about 7.5 cents suppose the average length of a sentence is 30 words, then the cost of human translation (or labeling) of 10 million sentences is about 22 million U.S. dollars. When considering the mutual translation of hundreds of languages, the cost of human labeling will reach hundreds of billions of dollars.

A notable progress in this direction was made by Turing Award winner Yann LeCun, who accelerated the research progress in self-supervised learning by learning supervised information from unlabeled data to improve the learning

ability on unlabeled data [14]. Organizations like DeepMind<sup>2</sup> and OpenAI<sup>3</sup> continue to evolve deep reinforcement learning algorithms in pursuit of autonomous decision-making and intelligence collaboration of multiple agents. Instead of handling a single modality, e.g., vision or speech, individually, multimodal learning can cope with more complex application settings by enabling multimodal understanding and learning inter-modal relationships from multimodal information sources. Typical applications include dish preparation videos aligned with recipe text steps [15], and lip-movement visual descriptions fused with speech signals to predict words [16].

The algorithmic route for cognitive understanding is still inconclusive, and it may take several years to come before the cognitive era advents. From the perspective of theoretical system, leading experts of deep learning have started to explore new forms of deep learning theoretical system; some researchers have started questioning existing fundamental theories such as back propagation and classical neural network models. Importantly, Geoffrey Hinton has proposed an alternative deep neural network (DNN) architecture, called the capsule network [17], attempting to address the challenge of small sample sizes. Although there is progress in capsule network research, each time, the research progress is not a superimposed evolution but rather an alternative of the original proposal. It has become a industry consensus in the past year that reinforcement learning may not achieve general intelligence. Instead, self-supervised learning has become a key path of interest for many experts, appearing frequently in academic venues like ICML and ICLR.

### B. Diversified Learning Styles for Complex Scenarios

The industry has been exploring multiple learning approaches, including reinforcement learning [13] and self-supervised learning [12] to make up the gap towards general intelligence in the face of complex task scenarios.

*Deep reinforcement learning* is expanding the boundaries to achieve multi-agent intelligence and autonomous decision-making in incomplete information gaming tasks, such as multiplayer chess and real-time strategy games [18]. OpenAI, Google, and Microsoft have accomplished multiple complex games (e.g., real-time strategy, Texas Hold'em, and Mahjong) and started attacking more practical scenarios (e.g., navigation of drones in groups). Deep reinforcement learning also gradually expand to applications of chip design, music arrangement and other professional fields that require higher level of domain knowledge and skills [19]. An example is Google researchers [20] used deep reinforcement learning to expedite and optimize the design of chip layout (with the aim of balancing power, performance, and area) in 2020; Tsinghua University [21] proposed deep reinforcement learning algorithms for generate accompaniment at real time based on the input music.

*Self-supervised learning* has become the most active research field. Google, Facebook and many other companies

have released algorithmic models based on self-supervised learning. These models can significantly reduce human intervention by mining supervised information from unlabeled data. In particular, in the field of natural language processing, Google BERT [22], Facebook RoBERTa<sup>4</sup>, and OpenAI GPT-3<sup>5</sup> have all made notable achievement. Currently, academia and industry are increasingly applying self-supervised learning to the field of computer vision (CV), and have made some progress in image processing. For example, University of Washington [23] used self-supervised learning to achieve front and back view separation of image backgrounds with pixel-level accuracy. The main limitation of current self-supervised models lies in the reliance on normalized and labeled data the models mainly construct and learn data features with the help of pre-trained models, rather than based on deep understanding of data and tasks.

### C. Vertical Integration with Domains

As the limitations of deep learning becomes evident, industry giants like Geoffrey Hinton has been promoting theoretic innovations in the field. Capsule network [17] tries to address the challenges of the high dependency on labeled data and explainability of deep learning models; this research, however, is still in the exploration stage. In 2020, Hinton's team [24] proposed a network detection mechanism, which significantly improved the attack detection rate in the field of machine learning security. Scholars [25] from University of Central Florida proposed a capsule routing method, which can query human and postures from videos according to conditions specified in sentences; this research, however, remains in the research phase.

Deep neural networks sped up in the integration with other techniques, as leading enterprises and universities seek innovations via integrating deep neural networks with knowledge graph [6] and traditional machine learning. Research in knowledge graphs aims to overcome the limitations of small-sample-size learning and insufficient understanding and reasoning ability of deep learning models without subverting the deep learning theory. Vertical domain-oriented knowledge graphs [26] have been applied in finance, medical, and judicial industries and significantly improved the intelligence in automatic knowledge association and acquisition in vertical industry applications. For example, the Ant Financial Knowledge Graph platform has been widely used in micro-loan, insurance claims, and smart finance within Ant and its partners<sup>6</sup>. Amazon has developed a Drug Repositioning Knowledge Graph (DRKG) to predict the possibility of combining drugs with disease targets, shorten the drug research and development cycle, and reduce costs<sup>7</sup>. Besides, new algorithmic forms emerge from integrating deep learning and traditional machine learning. A representative example is Bayesian deep

<sup>4</sup><https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/>

<sup>5</sup><https://openai.com/blog/gpt-3-apps/>

<sup>6</sup><https://alibaba-cloud.medium.com/ant-financial-applies-ai-in-financial-sector-6fe6e9186057>

<sup>7</sup><https://www.amazon.science/blog/amazon-web-services-open-sources-biological-knowledge-graph-to-fight-covid-19>

<sup>2</sup><https://deepmind.com/>

<sup>3</sup><https://openai.com>

learning [27], which has become a hot direction—Bayesian deep learning effectively uses prior knowledge to address the overfitting and small-sample-size issues, thus achieving superior performance to traditional deep learning methods. In particular, DeepMind proposed Bayesian RNN model [28], which performs significantly better than the traditional RNN model for graph annotation generation. Researchers from New York University and Samsung proposed a uncertainty representation method (SWAG) [29] for deep learning based on the Bayesian idea, which significantly improves the models generalization ability and performs well on multiple computer vision tasks, such as anomaly detection and calibration.

#### D. Pre-trained Models for General Intelligence

Pre-training model has embraced parameters of the trillion level, and the high training cost has almost become exclusive for the top players in the industry. In 2020, OpenAI released the GPT-3 model with 175 billion model parameters and a training cost of \$12 million, creating a barrier to the construction of pre-trained models that is difficult for small and medium-sized AI companies to meet. In 2021, Google released the SwitchTransformer model [30], again pushing the model parameters to a new high of 1.6 trillion. In addition, Microsoft announced a cooperation agreement with OpenAI to obtain the exclusive license of GPT-3 language model. These are significant events indicating that the future super large-scale pre-training models may be in the hands of a few giant enterprises.

Pre-training models have entered the age of general-purpose intelligence, with leading AI companies releasing generic pre-training models that are readily applicable to a variety of natural language processing tasks, eliminating the need for fine-tuning for different tasks. At present, leading general-purpose pre-training models like Google T5<sup>8</sup>, OpenAI GPT-3 can generally improve text comprehension and approach human-level accuracy on multiple benchmark tests including reading comprehension and Q&A tasks. Meanwhile, general-purpose pre-training models are accelerating into industrial applications. For example, OpenAI has released GPT-3 commercial application program interface (API) that provides Q&A, translation, and text generation services, and has been used by many companies such as Algolia (search service provider) and Reddit (social media platform).

#### E. Model Miniaturization for Model Efficiency

Efficiency improvement is becoming a new trend in of deep learning applications. Facing complex scenarios with massive data, the accuracy of deep learning models tends to increase with the growth in model complexity. Complex models generally incur omre computational load, making it difficult to deploy and run these models on devices with limited hardware capacity.

Model miniaturization aims to improve the efficiency of deep learning models. Techniques for model miniaturization

generally include knowledge distillation [31] pruning and quantization [8], which are becoming mature; and the mainstream models can reach tens of times of compression rate. For example, Amazon uses knowledge distillation to extract compressed model (Bort) from BERT [22]—the compressed model size is only 6% of BERTs size, and the inference speed is increased by seven times. Researchers from MIT and SJTU propose LiteTransformer [32], which combines quantization and pruning techniques to compress Transformer model by 95%. Industry wise, model compression has become a key capability of existing deep learning frameworks (e.g., TensorFlow<sup>9</sup>, PyTorch<sup>10</sup>, and TensorRT<sup>11</sup>), with leading AI companies (e.g., Facebook, Tencent, Google) and chip makers (e.g., Nvidia and Intel) accelerating the construction of model compression capabilities relying on their own algorithm technology and hardware chip advantages. These companies also provide algorithm compression tools such as pruning and quantization in their leading development frameworks such as TensorFlow, PyTorch and TensorRT, and carry out specific compression optimization for GPU, CPU and other hardware chips.

#### F. Applications-driven Revolution in Intelligence Computing

The broad application of deep learning promotes adoption of cloud computing for high performance computing. The training effect of deep learning still highly depends on computing resources and data quality, pushing the need for large-scale high-speed processing capabilities. Google predicted in 2013 that the computing power of the traditional CPU-based data center must be doubled if all users use the voice search function for 3 minutes a day. In fact, the expected growth in computing power demand has prompted Google to develop tensor processors for AI applications. As deep learning models becomes more complex and the training sample sizes continues growing, the demand for computing power is expect to keep increasing, placing higher demands on cloud-side computing performance.

The computing model is moving toward cloud-edge collaboration, and end-side computing power has become a new direction of explosion in research and development<sup>12</sup>. Edge scenarios (e.g., autonomous driving, industrial intelligence, and smart cities) represent a decentralized computing environment with specific requirements in power consumption, real-time, reliability, and security [33]. To accommodate these requirements, edge devices need to be deeply integrated with specific applications through multiple links (e.g., chip architecture, programming models, dedicated acceleration libraries, and software frameworks) to upgrade the full-stack capability of edge computing platforms. It is highly likely that scenario-based smart computing chips for industrial electronics, automotive electronics and traditional consumer electronics appli-

<sup>9</sup><https://www.tensorflow.org/>

<sup>10</sup><https://pytorch.org/>

<sup>11</sup><https://developer.nvidia.com/tensorrt>

<sup>12</sup><https://enterpriseproject.com/article/2020/12/8-edge-computing-trends-2021>

<sup>8</sup><http://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html>

cations will grow rapidly in the coming years, with annual compound growth rates of up to 100% or more in market volume, becoming the main driving force for the smart chip industry<sup>13</sup>.

### III. PROSPECTS FOR THE AI INDUSTRY

#### A. Transformation into Platform-based Development Models

AI development platforms are being developed at fast pace, as Chinas market for AI development reaches 140 million U.S. dollars in the first half of 2020, at a compound rate of increase of 30%<sup>14</sup>. The utilization rate of synchronous AI technical services has exceeded 1 trillion requests per day. Examples include Alibaba Cloud<sup>15</sup>(which processes 1 trillion requests per day and more than 1 billion images) Baidu Research<sup>16</sup> (which receive more than 1 trillion requests per day after opening more than 270 AI capabilities to the public). Meanwhile, the scale of AI ecosystems keeps increasing. For example, iFlytek<sup>17</sup> has 1.756 million developers in its development community, accumulating over 2.89 billion terminal devicesTencent AI<sup>18</sup> has served more than 1.2 billion customers worldwide, accumulating over 2 million users.

The potential for AI development platforms is unprecedentedly prosperous. Major cloud service providers like Amazon, Microsoft, and Google are leading the advancement, increasingly investing, and becoming more reliant on AI development platforms. Amazons SageMaker platform<sup>19</sup>, for example, has a doubled market share of the next two largest platforms, accounting for over 80% of TensorFlow load worldwide. Innovative startups, such as H2O.ai, DataRobot<sup>20</sup>, are emerging with unique experiences and expertise, many becoming sought-after partners within the industry.

In fact, competition has become inevitable as a crowded and varied market of AI development platforms emerges, centering around the technical *toolchain* of each platform. Currently, three development paths exist. First is institutionalization and systemization of the toolchain, where the construction of an intuitive and comprehensive toolchain is the key in this competition. The toolchains provide capabilities for data computation and processing, supporting the construction, deployment, and monitoring of various deep models, as observed from SageMaker Autopilot<sup>21</sup>, Google AutoML<sup>22</sup>, and Microsoft MLOps<sup>23</sup>. The second path is to open AI development platforms to support various frameworks, such as

TensorFlow, PyTorch, and MXNet<sup>24</sup>. The third path is to constantly optimize and improve decentralized computation with deep optimizations around specialization features and cloud compute architectures. For example, SageMaker achieves a scaling efficiency of 90% using 256 GPUs while supporting a substantial amount of dedicated chips and NPUs.

Technology service platforms are becoming mature, providing a variety of fundamental technical services in the realms of vision, voice, and natural language processing. Google, Microsoft, and Amazon are all serving as cloud service providers and have evolved in comprehensive platforms covering a variety of tasks. Other companies (e.g., iFlytek and Megvii<sup>25</sup>), in a relatively initial stage, are offering AI services. They mostly specialize in certain fields to maintain a technical advantage. For example, iFlytek focuses on text-to-speech translation while Megvii emphasize on vision processing. But it has become common for leading corporations to offer AI services in more than one field, thanks to reduced complexity and difficulty for a single service provider to offer services in more than one fields. Real-life applications play a key role in driving this mixture, as a single service is increasingly unable to fulfill dynamic demands in those applications.

Service platforms built on vertically integrated technologies represent another industrial trend. Vertical technology service platforms usually collect universally applied technologies to then duplicate and re-distribute. However, such platforms are still in infancy and have not reached a level of large-scale development. Currently, major platforms provide two forms of services: generating direct output of computational results (App program interface); providing readily applicable solutions. The former generally faces a vertical industry with mature applications that support downstream applications in using and integrating the information generated; it also has high prerequisite requirements for the application chain or faces limited market for expansion. As a result, this form of services may not form a strong competition against existing providers that have already integrated the industry chain.

AI technologies and its vertical industrial integration pose demands on the existing software, hardware, and technological solutions, inspiring novel solutions that can potentially address those challenges. Vertical service providers also promote rapid and seamless integration between intelligent technologies and its vertical industrial integration. The Apollo self-driving platform integrates radar, camera, high-resolution LIDAR and mapping, route planning services into an all-in-one solution.

#### B. Ecosystems with Open-source Development Frameworks

As the global AI industry grows up gradually, open-source development frameworks present a venue where multiple subjects like chips and application development converge, and the ecology built around the open-source development frameworks of companies like Google and Facebook has taken initial shape. The AI industry has not only transformed the

<sup>13</sup><https://www.gsma.com/membership/wp-content/uploads/2020/11/5G-Digital-World-Built-on-Chips.pdf>

<sup>14</sup><https://www.chinadaily.com.cn/a/201912/08/WS5decae0ca310cf3e3557ca23.html>

<sup>15</sup><https://www.googleadservices.com>

<sup>16</sup><http://research.baidu.com/>

<sup>17</sup><http://www.iflytek.com/en/>

<sup>18</sup><https://ai.tencent.com/ailab/en/about/>

<sup>19</sup><https://aws.amazon.com/sagemaker/>

<sup>20</sup><https://www.datarobot.com/>

<sup>21</sup><https://aws.amazon.com/sagemaker/autopilot/>

<sup>22</sup><https://cloud.google.com/automl>

<sup>23</sup><https://azure.microsoft.com/en-au/services/machine-learning/mlops/>

<sup>24</sup><https://mxnet.apache.org/>

<sup>25</sup><https://en.megvii.com/>

information industry with respect to product form and operation mode but also affected the production and operation efficiency of traditional industries. Therefore, players in the AI industry almost cover all information industry and traditional industry enterprises. Cloud computing, chips, ICT equipment, Internet, intelligent technology services, vertical industry AI companies, and traditional industry enterprises are constantly exploring industrial ecologies based on their own advantages.

Leading AI companies that build open-source development framework aim to form a dominance in both application interfaces and hardware adaptation. On the one hand, development frameworks like TensorFlow and PyTorch continue to improve their high-level language interfaces and develop algorithm models and smart applications based on their high-level language interfaces, forming deep binding between the two. Since switching to another framework will increase secondary R&D cost and likely affect model performance, in the long run, many applications will rely on the original development frameworks for training and reasoning, forming the industry's default *de facto* interface standard. On the other hand, leading AI companies with large market shares are building standardized hardware interfaces on top of their frameworks and smart chip adaptations, which will further enhance their voice in the smart-chip industry. For example, by building a unified compiled intermediate representation language (IR), Google drives hardware manufacturers to actively adapt to its TensorFlow framework, shifting from smart chip-led adaptation to IR-led adaptation.

As the main body of the AI industry cuts in with its own advantages, the industry has preliminarily formed four ecological models.

1) *Cloud-based AI infrastructure*: AI is fully integrated into cloud service systems, and cloud service vendors are actively building an AI infrastructure ecosystem. Manufacturers represented by Amazon, Microsoft, and Google have laid out full-stack AI support systems, covering dedicated AI hardware, open-source development tools, research platforms, technical services, and industrial solutions. In 2020, Microsoft established an intelligent supercomputing centre and launched the industry cloud solution (Microsoft Cloud<sup>26</sup>) for Healthcare. Microsoft also cooperated with smart healthcare and medical institutions to apply AI technology to precision medical services.

2) *Application-specialized platforms and solutions*: AI companies are increasingly using vision, voice and other technologies to create of ecosystems with vertical industrial service platforms and solutions, along with industrial application of smart technologies. Typical domain-facing platforms include Yitu<sup>27</sup> (Medical) and Baidu Apollo Open Platform<sup>28</sup> (Automatic Driving).

3) *Traditional industry transformation*: Traditional industries are cutting in with industry experience and ability to solve

practical problems. They actively build multi-dimensional innovation ecology covering scientific research, results transformation, and industry cultivation. Leading companies in traditional industries are deploying disruptive and cutting-edge technologies to consolidate their leading positions and advantages in their own industries. For example, Volkswagen Audi established the Audi.JKU deep learning centre<sup>29</sup>, focusing on autopilot; it also increased efforts on developing advanced drivers Auxiliary systems and released automatic driving data sets to promote commercial and academic research.

4) *Software-hardware collaboration*: Hardware manufacturers are using chip design and machine integration as the entry point to construct software-hardware collaborative ecosystems. For example, Nvidia, Intel, Huawei, and Inspur are all actively enriching their compilers, programming frameworks, tools, and supporting software around their chips and computing equipment. By expanding the developer community and industry partners, they aim to build an industrial ecosystem for software and hardware synergy.

#### IV. CONCLUSION

In this paper, we review trends and state-of-the-art development of the AI industry and related technologies, followed by remarking on prospects for the AI industry. This paper will facilitate researchers and practitioners to gain a systematic understanding of the industry. It also serves as a point of reference for stakeholders in related industries. While this paper does not cover much about the ethical aspect of industrial AI, we believe in the necessity of promoting democratization of AI in academia and for small companies in the industry.

#### REFERENCES

- [1] G. Serin, B. Sener, A. Ozbayoglu, and H. Unver, "Review of tool condition monitoring in machining and opportunities for deep learning," *The International Journal of Advanced Manufacturing Technology*, pp. 1–22, 2020.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [3] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.
- [4] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [5] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *arXiv preprint arXiv:2001.06937*, 2020.
- [6] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Systems with Applications*, vol. 141, p. 112948, 2020.
- [7] S. Zhao, B. Li, P. Xu, and K. Keutzer, "Multi-source domain adaptation in the deep learning era: A systematic survey," *arXiv preprint arXiv:2002.12169*, 2020.
- [8] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PloS one*, vol. 14, no. 11, p. e0224365, 2019.
- [9] W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: A survey," *IEEE Access*, vol. 7, pp. 63 373–63 394, 2019.

<sup>26</sup><https://azure.microsoft.com/en-au/>

<sup>27</sup><https://www.yitutech.com/en>

<sup>28</sup><https://apollo.auto/>

<sup>29</sup><https://www.audi-mediacycenter.com/en/press-releases/audi-and-johannes-kepler-university-of-linz-to-establish-center-for-artificial-intelligence-9044>

- [10] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.
- [11] T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, J. Kautz, and B. Catanzaro, "Few-shot video-to-video synthesis," *arXiv preprint arXiv:1910.12713*, 2019.
- [12] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [13] T. M. Moerland, J. Broekens, and C. M. Jonker, "Model-based reinforcement learning: A survey," *arXiv preprint arXiv:2006.16712*, 2020.
- [14] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," *arXiv preprint arXiv:2103.03230*, 2021.
- [15] J. Malmaud, J. Huang, V. Rathod, N. Johnston, A. Rabinovich, and K. Murphy, "What's cookin'? interpreting cooking videos using text, speech and vision," *arXiv preprint arXiv:1503.01558*, 2015.
- [16] R. Gu, S.-X. Zhang, Y. Xu, L. Chen, Y. Zou, and D. Yu, "Multi-modal multi-channel target speech separation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 3, pp. 530–541, 2020.
- [17] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *arXiv preprint arXiv:1710.09829*, 2017.
- [18] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.
- [19] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications," *IEEE transactions on cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020.
- [20] A. Mirhoseini, A. Goldie, M. Yazgan, J. Jiang, E. Songhori, S. Wang, Y.-J. Lee, E. Johnson, O. Pathak, S. Bae *et al.*, "Chip placement with deep reinforcement learning," *arXiv preprint arXiv:2004.10746*, 2020.
- [21] N. Jiang, S. Jin, Z. Duan, and C. Zhang, "RI-duet: Online music accompaniment generation using deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 710–718.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [23] I. Katircioglu, H. Rhodin, V. Constantin, J. Spörri, M. Salzmann, and P. Fua, "Self-supervised segmentation via background inpainting," *arXiv preprint arXiv:2011.05626*, 2020.
- [24] Y. Qin, N. Frosst, C. Raffel, G. Cottrell, and G. Hinton, "Deflecting adversarial attacks," *arXiv preprint arXiv:2002.07405*, 2020.
- [25] B. McIntosh, K. Duarte, Y. S. Rawat, and M. Shah, "Visual-textual capsule routing for text-based video segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9942–9951.
- [26] B. Abu-Salih, "Domain-specific knowledge graphs: A survey," *Journal of Network and Computer Applications*, vol. 185, p. 103076, 2021.
- [27] H. Wang and D.-Y. Yeung, "Towards bayesian deep learning: A framework and some existing methods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3395–3408, 2016.
- [28] M. Dusenberry, G. Jerfel, Y. Wen, Y. Ma, J. Snoek, K. Heller, B. Lakshminarayanan, and D. Tran, "Efficient and scalable bayesian neural nets with rank-1 factors," in *International conference on machine learning*. PMLR, 2020, pp. 2782–2792.
- [29] W. J. Maddox, P. Izmailov, T. Garipov, D. P. Vetrov, and A. G. Wilson, "A simple baseline for bayesian uncertainty in deep learning," *Advances in Neural Information Processing Systems*, vol. 32, pp. 13 153–13 164, 2019.
- [30] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *arXiv preprint arXiv:2101.03961*, 2021.
- [31] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1789–1819, 2021.
- [32] Z. Wu, Z. Liu, J. Lin, Y. Lin, and S. Han, "Lite transformer with long-short range attention," *arXiv preprint arXiv:2004.11886*, 2020.
- [33] J. Pan and J. McElhannon, "Future edge cloud and edge computing for internet of things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2017.