A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development

Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, *Fellow, IEEE*, and Yang Tang, *Senior Member, IEEE*

***Abstract*—ChatGPT, an artificial intelligence generated content (AIGC) model developed by OpenAI, has attracted worldwide attention for its capability of dealing with challenging language understanding and generation tasks in the form of conversations. This paper briefly provides an overview on the history, status quo and potential future development of ChatGPT, helping to provide an entry point to think about ChatGPT. Specifically, from the limited open-accessed resources, we conclude the core techniques of ChatGPT, mainly including large-scale language models, in-context learning, reinforcement learning from human feedback and the key technical steps for developing ChatGPT. We further analyze the pros and cons of ChatGPT and we rethink the duality of ChatGPT in various fields. Although it has been widely acknowledged that ChatGPT brings plenty of opportunities for various fields, mankind should still treat and use ChatGPT properly to avoid the potential threat, e.g., academic integrity and safety challenge. Finally, we discuss several open problems as the potential development of ChatGPT.**

***Index Terms*—AIGC, ChatGPT, GPT-3, GPT-4, human feedback, large language models.**

I. Introduction

ARTIFICIAL intelligence generated content (AIGC), which is one of the most fascinating frontier technology, refers to that users can use AI to create contents (e.g., images, text, and videos) automatically according to their personalized requirements [1]−[3]. With the iterative development of AI algorithms and network structures [4], [5], significant progress has been made in AIGC. Generative adversarial network (GAN) [6], [7], contrastive language-image pre-training (CLIP) [8], diffusion model [9], [10] and multimodal generation [11], [12] are core technologies for various fields of AIGC so that contents with high quality can be generated automatically. Moreover, with the advancement of GPU and the increase in computational power, large scale deep networks with huge amounts of parameters are trained so that more information can be learned to deal with general downstream tasks with better performance [13]. Based on the above technological developments, in the year of 2022, various of AIGC products were developed and iterated by the world's leading technology companies, for example DALL-E-2 of OpenAI [14] can generate images of high-quality with giving specific descriptions and Meta proposes Make-A-Video [15] to directly translate texts to videos. In the end of 2022, OpenAI released the public version of ChatGPT, which further attracts worldwide attention on perfectly responding to any human requests described in natural language. ChatGPT has surpassed 100 million monthly active users till the end of January 2023, just two months after its launch, according to the report of UBS[1]. In this paper, we carry out several in-depth thoughts and discussions on ChatGPT.

ChatGPT is an intelligent chatting robot which is able to provide a detailed response according to an instruction in a prompt. As a member of AIGC, ChatGPT has shown powerful functions on various language understanding and generation tasks such as multilingual machine translation, code debugging, stroy writing, admitting mistakes and even rejecting inappropriate requests according to the official statement. Unlike previous chatting robots, ChatGPT is able to remember what the user has said earlier in the conversation which helps for continuous dialogue [16]. In March 2023, with the publication of GPT-4 created by OpenAI [17], ChatGPT also has enjoyed a strong update with further functions. Specifically, users now can input texts and visual images in parallel, thus more challenging multimodal tasks can be completed such as image captioning, chart reasoning, paper summarizing[2] (as shown in Fig. 1).

ChatGPT, with so many strong functions, is an integration of multiple technologies such as deep learning, unsupervised learning, instruction fine-tuning, multi-task learning, in-context learning and reinforcement learning. It is built upon the initial GPT (Generative pre-trained Transformer) model, which has been iteratively updated from GPT-1 to GPT-4 (as shown in Table I). GPT-1 [18], developed in 2018, is firstly

Manuscript received March 25, 2023; revised April 13, 2023; accepted April 15, 2023. This work was supported by National Key Research and Development Program of China (2021YFB1714300), National Natural Science Foundation of China (62293502, 61831022, 61976211) and Youth Innovation Promotion Association CAS. Recommended by Associate Editor Xin Luo. *(Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, and Kang Liu contributed equally to this work. Corresponding authors: Qing-Long Han and Yang Tang.)*

Citation: T. Y. Wu, S. Z. He, J. P. Liu, S. Q. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of ChatGPT: The history, status quo and potential future development," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 5, pp. 1122–1136, May 2023.

y20200093@mail.ecust.edu.cn.

shizhu.he@nlpr.ia.ac.cn.

jingpingliu@ecust.edu.cn.

siqisun@fudan.edu.cn.

kliu@nlpr.ia.ac.cn.

qhan@swin.edu.au.

yangtang@ecust.edu.cn.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2023.123618

[1] https://archive.is/XRl0R

[2] https://openai.com/research/gpt-4

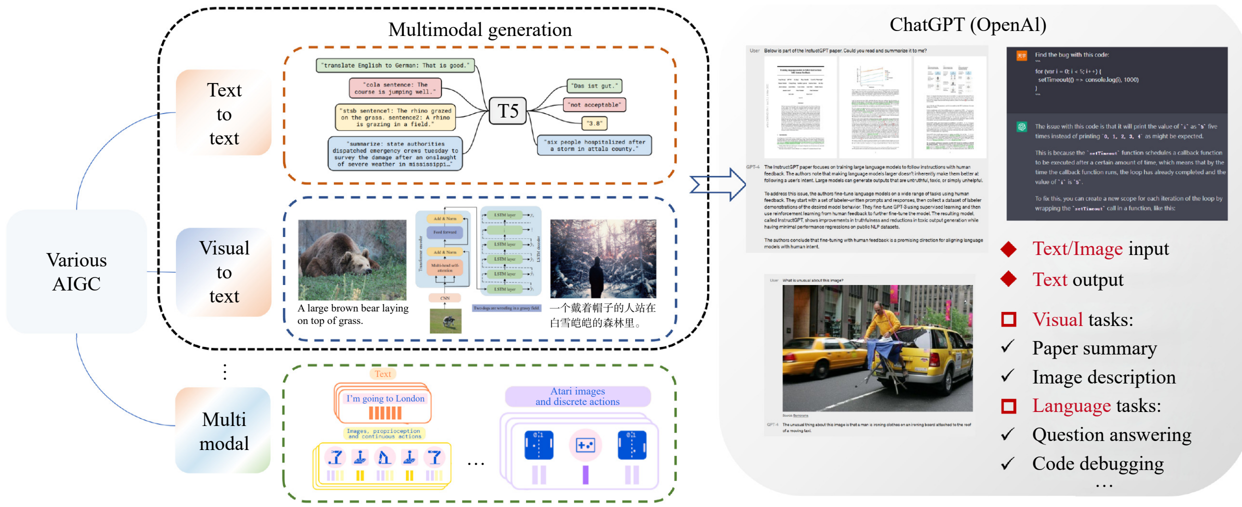


Fig. 1. Strong function of AIGC, including text-to-text generation, visual-to-text generation and further multimodal generation. For ChatGPT of OpenAI, which combines text and visual images as input, can handle various of language and visual tasks. Cases are adapted from [1], [3], [11], [17].

TABLE I
COMPARATIVE ANALYSIS OF GPT-1, GPT-2, GPT-3 AND GPT-4

	GPT-1	GPT-2	GPT-3	GPT-4
Released date	June 2018	February 2019	May 2020	March 2023
Model parameters	117 million 12 layers 768 dimensions	1.5 billion 48 layers 1600 dimensions	175 billion 96 layers 12 888 dimensions	Unpublished
Context window	512 tokens	1024 tokens	2048 tokens	8195 tokens
Pre-training data size	About 5 GB	40 GB	45 TB	Unpublished
Source of data	BooksCorpus, Wikipedia	WebText	Common Crawl, etc.	Unpublished
Learning target	Unsupervised learning	Multi-task learning	In-context learning	Multimodal learning

dedicated to train a generative language model based on a Transformer framework via unsupervised learning [19]–[21], and the pretrained model is further fine-tuned on downstream tasks. GPT-2, developed in 2019 [22], mainly introduces the idea of multi-task learning [23] with more network parameters and data than GPT for training, so that the pretrained generative language model can be generalized to most of the supervised subtasks without further fine-tuning. To further improve the model performance on few-shot or zero-shot [24] settings, GPT-3 [25] combines meta-learning [26], [27] with in-context learning [28] so that the generalization ability of the model has been greatly improved with surpassing most of the existing methods on various downstream tasks. Moreover, the parameter scale of GPT-3 has increased by 100 times than GPT-2 and it is the first language model to surpass a parameter scale of 100 billion. When it comes to the pilot version of ChatGPT (InstructGPT, also known as one of the derivative version of GPT3.5 series models), the researchers use reinforcement learning with human feedback (RLHF) to incrementally train the GPT-3 model [29] so that the model can better follow and align with the user's intent. Finally, when it comes to GPT-4, a large multimodal model with accepting image and text inputs and emitting text outputs, ChatGPT exhibits human-level performance on arious professional and academic benchmarks [17].

In addition to the quick update on techniques, the enlargement of model capacity and the number of data for pre-training further helps the model to better understand the meaning and intent behind a given piece of text. The parameters and the data of GPT-1 only reach 117 million and 5 GB, respectively, while in GPT-3 the number increases to 175 billion and 45 TB. Even if the details of GPT-4 have not been publicly released, it is expected that the parameters and data may still have a huge increase. In this way, the success of ChatGPT relies on multi aspects of supports such as financial backing, computational power, data resources. Till now, ChatGPT/GPT-4 has shown promising performance as a general purpose multimodal task-solver [30], which has been widely used in the field of we-media, data analysis, etc. However, the strong creativity of ChatGPT is a double-edged sword, especially in the filed of education and science [31], [32]. How to prevent the cheating or plagiarism and how to avoid the abuse of strong ChatGPT are still important topics for further discussion.

The aim of this paper is to provide a brief overview on the history, status quo and potential future development of ChatGPT, helping to provide an entry point to think about ChatGPT. Main contributions are as follows:

1) Core techniques of ChatGPT, mainly including pre-trained large-scale language models, in-context learning, reinforcement learning from human feedback and the key technical steps for developing ChatGPT are specified.

2) The pros and cons of ChatGPT and the duality of ChatGPT in various field are provided.

3) Several open problems on potential research trends of

ChatGPT are discussed.

In Section II, we delve into the core techniques of ChatGPT. In Section III, we discuss the advantages and disadvantages of the ChapGPT. In Section IV, we present the potential impact of ChatGPT across various social fields. In Section V, we consider the future research trends of ChatGPT. Finally, a simple conclusion is conducted in Section VI.

II. Important Technologies Behind the ChatGPT

Although the technical details (including architecture, hardware, dataset construction and training method, or similar) of GPT-4 has not been shared, the main techniques of GPT-4 are still similar to GPT-3/GPT-3.5 [17]. In this way, we can only analyze it based on the public information and its twin model, InstructGPT [29]. First, we introduce the pre-trained language model, which is the foundation technology of ChatGPT. Then, we introduce the in-context learning that models general tasks with a self-learning paradigm, we also append the novel technologies of chain-of-thought prompting and instruction fine-tuning into this subsection. Next, we introduce the reinforcement learning from human feedback, which can continuously optimize the dialogue model. Finally, we introduce the three main steps in the implementation of ChatGPT/InstructGPT with public information: supervised fine-tuning (SFT), reward modeling (RM) and dialogue-oriented RLHF.

A. Pre-Trained Language Model

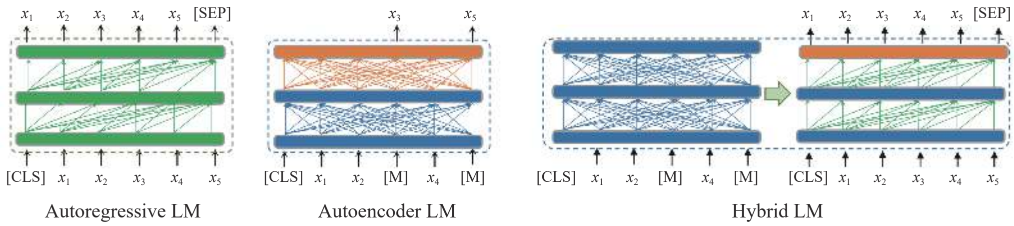
Language models are statistical models that describe the probability distribution of natural language [33]. It is dedicated to estimating the probability of a given sentence (e.g., $P(S) = P(w_1, w_2, \ldots, w_n)$ used to compute the probability of the sentence S containing n words), or the probability of generating other contents given a part of the sentence (e.g., $P(w_i, ..w_n | w_1, \ldots, w_{i-1})$ used to compute the probability of predicting the content of the next part given the content of the previous part of the sentence), which is the core task of natural language processing and can be used on almost all downstream NLP tasks.

Different modeling methods of statistical language models denote the technical level of natural language processing. In the early n-gram language model (N-gram LM), the conditional probability is estimated by the frequency statistics of n-grams $P(w_i | w_{i-1}, \ldots, w_{i-(n-1)}) = \frac{C(w_{i-1}, \ldots, w_{i-(n-1)})}{C(w_{i-1}, \ldots, w_{i-(n-1)}, w_i)}$. It can only consider a fixed length (window size) of word sequences based on the Markov assumption, and the probability estimation of language model is inaccurate due to the cures of dimensionality with symbol combinations. N-gram LM drives the development of information retrieval technology based on keyword search and document relevance computation. The neural language model (NLM) [34], [35] started in 2010 has driven researchers to model complex natural language and multiple tasks [36], [37] with a wide range of deep neural network structures such as multilayer perceptron (MLP), convolutional neural network (CNN) and recurrent neural network (RNN), and thus formed representative static language models such as word2vec [38], [39] and RNNLM [40], [41]. NLM utilizes low-dimensional embedding to represent words and their composition, and realizes the prediction of conditional probability through the calculation of neural network.

Since 2018, the pre-trained language models (PLM), which utilize self-supervised learning over raw large-scale texts, have received more and more attention [42]. And it promotes the birth and development of the two-stage learning paradigm of pre-training and fine-tuning. For example, the relevant models of ELMo [43], BERT [44] and GPT-3 [25] have won the best paper award of NAACL 2018, NAACL 2019 and NeurIPS 2020, respectively. PLM learns general language models on large-scale texts based on self-learning tasks such as masked word prediction [44], sequence recognition of sentences, text filling in the blank [45], and text generation [2]. It not only improves the semantic description of words from static representation to context-aware dynamic representation, but also provides a unified modeling framework for NLP tasks.

At present, there are three typical model structures of PLM, autoregressive LM, autoencoding LM, and hybrid LM (as shown in Fig. 2). Their representative models are GPT [18], BERT [44] and T5 [1], respectively. Autoregressive LM is a standard language model, which adopts decoder-only manner of language modeling with one-way language encoding-decoding and token-by-token prediction of words. Autoencoding LM randomly masks the words in the sentence, and utilizing bidirectional encoding, and then predicts the masked words based on the context encoding information (e.g., take the input "[CLS]$x_1 x_2$[M]$x_3 x_4$[M]" and predict the words "x_3" and "x_5" on the masked position marked by "[M]"). Hybrid LM combines the two methods above. After randomly masking the words in the sentence and bidirectional encoding, input the previous text in one direction and predict the subsequent words step by step (e.g., take the input "[CLS]$x_1 x_2$[M]$x_3 x_4$[M]" and "[CLS]$x_1 x_2 x_3 x_4 x_5$" and gradually output "$x_1 x_2 x_3 x_4 x_5$ [SEP]"). Since BERT released by Google has refreshed the best record of 11 NLP tasks at the beginning, and GPT-2's performance on typical NLP tasks is not better than BERT, BERT and its autoencoding pre-training methods have been followed by the vast majority of academia and industry [46]–[52].

Comparatively, with the use of the best neural network structure Transformer [53] at present, OpenAI still adheres to the autoregressive method and continues to release GPT-1 [18], GPT-2 [22], GPT-3 [25] and GPT-4 [17] (as shown in Table I). GPT-1 first learns a general language model on unlabeled raw texts, and then fine-tunes it according to specific tasks. Although GPT-1 has some effect on NLP tasks that have not been fine-tuned, its generalization ability is far lower than fine-tuned models on supervised tasks. GPT-2 does not carry out different model framework with GPT-1 but uses larger network parameters and learned on more datasets. GPT-3 follows the structure of GPT-2, but has made great improvement in model capacity. Surprisingly, the GPT-3 shows excellent performance in many text generation tasks such as machine translation, question answering, dialogue, reading comprehension and story generation. In fact, it is difficult to distinguish the text generated by GPT-3 and written by humans. Moreover, GPT-3 easily supports zero-shot and few-



	Autoregressive LM	Autoencoder LM	Hybrid LM
Encoding/Decoding strategy	Unidirectional encoder + step-by-step decoding	Bidirectional encoder + masked based decoding	Sequence to sequence
Input-output example	[CLS]$x_1x_2x_3x_4x_5$ $\rightarrow x_1x_2x_3x_4x_5$[SEP]	[CLS]x_1x_2[M]x_4[M] $\rightarrow x_3x_5$	[CLS]x_1x_2[M]x_4[M] + [CLS]$x_1x_2x_3x_4x_5$ $\rightarrow x_1x_2x_3x_4x_5$[SEP]
Learning paradigm	Pre-training (zero-shot)	Pre-training + Fine-tuning	Pre-training + Fine-tuning
Support tasks	NLU + NLG	NLU	NLU + NLG
Representative model	GPT	BERT	BERT/T5

Fig. 2. There three typical types of PLM: autoregressive LM, autoencoder LM and hybrid LM. Their working mechanisms are listed and compared in this table.

shot learning scenarios [54]. Since then, the pre-trained model has entered the era of large-scale parameters (also known as large language model [55] or foundation model [56]). Building on the success of GPT-3, OpenAI has continued to develop multiple GPT-3.5 series models such as Code-davinci-001, Text-davinci-001, Code-davinci-002, and Text-davinci-002 through technical improvements such as code-based pre-training, instruction fine-tuning, and reinforcement learning from human feedback[3]. OpenAI provides Playground and APIs for users to use these advanced models. Till now, when it comes to GPT-4 [17], which is a large multi-modal model as the latest milestone in OpenAI's effort in scaling up deep learning, GPT-4 shows stronger capability to handle much more nuanced instructions than GPT-3 based models.

The development of artificial intelligence technology shows that large models can learn more complex and higher-order features/patterns from raw data (texts for large language models, text-image for large multimodal models), so as to exhibit stronger capabilities in understanding and generating data. By learning all kinds of abstract knowledge from the raw data, large-scale pre-trained language models have better generality and generalization. In addition, to better support general task processing capabilities (realizing the general artificial intelligence), the autoregressive language model adopted by GPT-3 and its subsequent GPT3.5 series models has more advantages, and it can directly utilize natural language to describe different tasks in different fields. Furthermore, researchers are looking forward to the transparency of GPT-4.

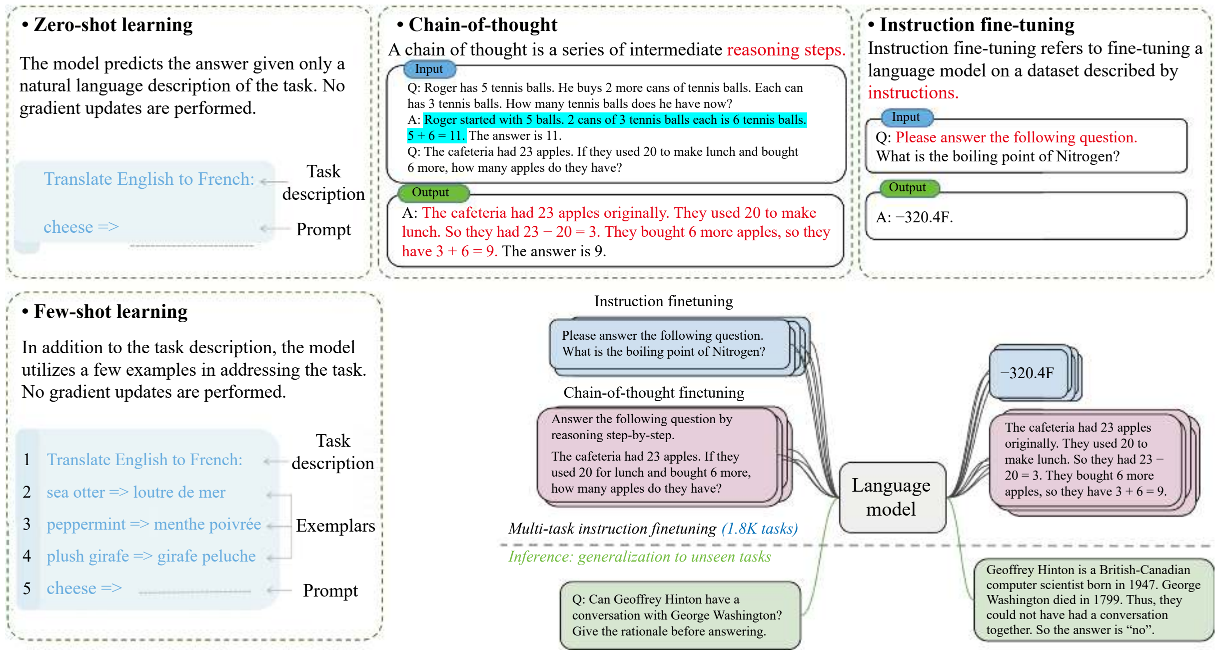
B. In-Context Learning

GPT-3 and the following GPT-3.5 serial models are the foundation of many powerful capabilities of ChatGPT. Among them, the in-context learning (ICL) introduced by GPT-3 plays a crucial role. As a meta-learning method that contains an internal loop, ICL can model more context information to solve specific tasks, which can not only improve the effects of various tasks, but also better deal with zero-shot and few-shot learning scenarios. With the support of ICL, GPT-3.5 series models can achieve good results without any training and fine-tuning of NLP tasks, and even achieved very shocking results in some logical and creative tasks such as article generation and code writing. Next, we will provide a detailed introduction to ICL and the technologies of chain-of-thought and instruction fine-tuning that promote the success of large models.

1) In-Context Learning: ICL has become a new paradigm of natural language processing (NLP) [28]. ICL can append a few exemplars to the context, which allows the model to learn and complete tasks by imitation. For example, as shown in Fig. 3 (left part), in order to translate the English phrase "cheese" into French, the task description and related exemplars are concatenated and input into the PLM as the context, and the model is allowed to generate the France phrase autonomously. ICL estimates the likelihood of potential answers over a trained language model. The core idea of ICL is to learning to complete tasks with analogies. Supervised learning or fine-tuning requires the use of backward gradients to update model parameters in the training phase. Unlike this, ICL does not need parameter updates, it directly performs analogical learning and task predictions over pre-trained language models. ICL expects to learn the hidden patterns in the demonstrations and make correct predictions. By adapting ICL, the zero-shot and few-shot learning capabilities are formed by directly describing the task or appending a small number of exemplars into the content as prompts.

2) Chain-of-Thought: Recently, chain-of-thought (CoT) prompting is proposed to further improve the ability to solve complex tasks such as answering arithmetic, commonsense, and logical reasoning questions [57]. CoT is dedicated to constructing a series of intermediate steps to simulate the thinking process of humans in completing complex tasks [58]. With CoT, LLM such as GPT-3 can be used to generate the reasoning steps and answers at the same time. For example, as

[3] https://yaofu.notion.site/GPT-3-5-360081d91ec245f29029d37b54573756



Flan-T5 [64] finetune various language models on 1.8K tasks phrased as instructions.

Fig. 3. Examples of different types in ICL, mainly including zero/few-shot learning, chain-of-thought and instruction fine-tuning (adapted from [64]).

shown in Fig. 3 (middle part), in order to answer the question "The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?", CoT is driven to imitate the previous example (the output "Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5+6 = 11. The answer is 11." for the input "Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can have 3 tennis balls. How many tennis balls does he have now?") and generate a description of the reasoning process "The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23–20 = 3.They bought 6 more apples, so they have 3+6 = 9." and the answer description "The answer is 9." in the following. Subsequently, researchers began to improve the reasoning of large language models from multiple different research perspectives, such as automatically generating CoT prompts (zero-shot reasoner) [59], generating multiple associated simple sub-tasks to complete a complex task [60], [61], building multiple reasoning chains to complete target tasks in combination with consistent learning [62], and select a better answer by self-verification [63].

3) Instruction Fine-Tuning: Moreover, in order to improve the task generalization ability of the large language model and deal with new tasks, researchers began to explore instruction fine-tuning (IFT) [64], [65]. That is, IFT describes all NLP tasks using natural language instructions and fine-tune large language models, so as to achieve the general ability of understanding and processing instructions. For example, as shown in Fig. 3 (right part), IFT places the natural language instructions "Please answer the following question." in front of the question "What is the boiling point of Nitrogen?". IFT unifies almost all tasks into the same text-to-text form based on the generative pre-training model. For example, as a representative model of the IFT model, FLAN [66] uses 137B parameters to continually train the pre-trained language model, and adjusts model parameters on more than 60 NLP datasets through natural language instructions. Flan-T5 [64] greatly improves the performance of the large language model on specific NLP tasks and the generalization ability for new tasks.

C. Reinforcement Learning From Human Feedback

Reinforcement learning (RL) mainly focuses on learning the optimal policy to maximize the desired reward or reach specific targets through the interactions between the agent and the environment [67]. Reinforcement learning has shown strong capacities on tasks with large action spaces, e.g., gaming [68]–[70], robotics control [71]–[73], molecular optimization [74], [75] and other fields [76], [77]. The general paradigm of reinforcement learning is a Markov decision process (MDP). An MDP can be defined as a tuple $M = (S, A, R, T, p)$, where S is the state space, A is the set of actions, $R : S \times A \to \mathbb{R}$ is the reward function, $T(s_{t+1}|s_t, a_t)$ is the transition probability from state s_t to state s_{t+1} when taking action a_t, $p(s_0)$ is the initial state distribution. Moreover, a policy $\pi(a|s)$ is a function which computes the corresponding conditional probability. In this way, the goal of MDP is to learn a optimal policy π^* and the cumulative reward is maximized within an episode of length L:

$$\pi^* = \arg\max_{\pi \in \prod} \mathbb{E}_{s\sim p(s_0)}[\mathcal{R}(s)] \tag{1}$$

where $\prod$ is the set of all policies, and the return of the state $\mathcal{R}(s)$ can be calculated as

$$\mathcal{R}(s) = \mathbb{E}_{a_t\sim\pi(a_t|s_t), s_{t+1}\sim T(s_{t+1}|s_t,a_t)}\Big[\sum_{t=0}^{L} R(s_t, a_t, s_{t+1})\Big]. \tag{2}$$

Up to now, various RL algorithms have been developed for different situations [78], e.g., TRPO [79], SAC [80], PPO [81], TD3 [82], and REDQ [83].

One crucial factor for training a successful reinforcement

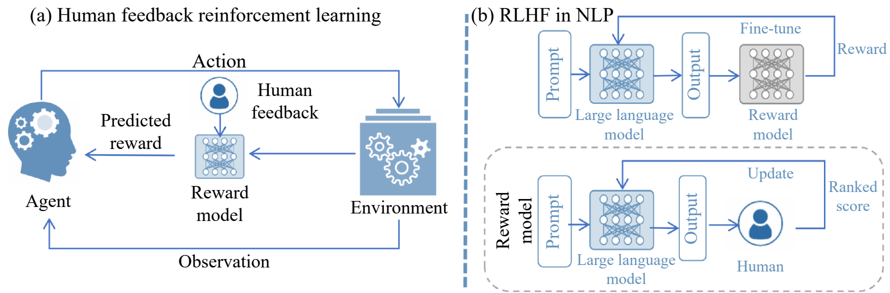


Fig. 4. Schematic illustration of (a) reinforcement learning from human feedback (RLHF) and (b) reinforcement learning from human feedback in NLP.

learning agent is a well-specified reward function. However, for tasks (e.g., table cleaning or clothes folding) of which goals are complex and poorly-defined, it is difficult to construct a precise reward function to evaluate whether the task is well finished with limited sensor information [84]. In this way, to further avoid the misalignment between human preferences and the objectives of RL agents, and to accelerate the efficiency when learning from scratch, human intuition or human expertise are further considered for knowledge transferring [85]. With the quick development of deep reinforcement learning, this technique has attracted lots of attention [86], [87] and can be concluded as reinforcement learning with human feedback (RLHF) [29] (as shown in Fig. 4(a)), which has also been applied in ChatGPT.

In fact, using human feedback directly when training the agent is quite expensive since massive amount of experience is needed to be evaluated manually. Hence, a reward model is trained to replace such works and rewards of human preferences can be provided in an economical way with offering several human data for training [84].

The process of training the reward model can be mainly divided into three steps: 1) the agent interacts with the environment under the current policy π to collect a series of trajectories of states, actions and rewards $\{s_1, a_1, r_1, \ldots, s_t, a_t, r_t\}$, 2) pairs of segments are selected from the trajectories generated in step 1, and thees segments are sent to humans for comparison and evaluation, 3) the parameters of the reward model is updated via supervised learning to fit human preferences. In this way, RLHF can be applied for more complex or customized tasks with a well-specified reward model.

In the natural language process, aligning the pre-trained large-scale language models (LMs) with human preference is greatly decided whether the models can generate truly "good" texts. Thus, RLHF has been positively applied to fine-tuning language models in most domains in natural language process (NLP) such as dialogue [88], [89], translation [90], [91], story generation [92], evidence extraction [93], [94], and semantic parsing [95]. Note that in GPT-4, an additional safety reward signal has been incorporated to reduce harmful outputs, helping to judging the safety boundaries for risk relieving. With applying RLHF as a low-tax alignment technique, language models become much more robust to handle real-world variations and nuances in that language. However, there still exist limitations and challenges deserve to be addressed, including the demand for high-quality human feedback, the time and cost required to obtain this feedback, and the effectiveness of reinforcement learning algorithms.

D. Key Technical Steps for Developing ChatGPT/InstructGPT

The technical details of ChatGPT and GPT-4 have not been shared, and the official[4] states that its implementation is similar to InstructGPT [29], which is a sibling model to ChatGPT, yet distinguished by the data collection and the pre-trained backbone. Hence, we next describe the technical steps of InstructGPT, mainly including supervised fine-tuning (SFT), reward modeling (RM), and reinforcement learning (RL), as depicted in Fig. 5.

1) SFT Model: In InstructGPT, this is a supervised policy model that is fine-tuned on GPT-3 [25]. The prompt is the input, and the response is the output. Note that the backbone of ChatGPT is a pre-trained model from the GPT-3.5 series.

Since SFT is a supervised model, it requires labeled data for training. Data is gathered from two different sources. First, some of the data is sampled from the OpenAI API of the earlier InstructGPT version, which was trained with a subset of demonstration data. Second, others are provided by labeler, which includes three types of prompts: plain (any arbitrary task), few-shot (an instruction with multiple query/response pairs), and user-based prompts (specific use cases that were requested for application in OpenAI). For each natural language prompt, the task is accompanied directly by an instruction (e.g., "Tell me about..."), but also indirectly through few-shot examples (e.g., giving two examples of a story, and prompting the model to write another story about the same topic) or implicit continuation (e.g., giving the start of a story, and asking the model to finish it). As a result, the SFT dataset has about 13k training prompts.

2) RM Model: This is a reward model that takes a pair of prompt and response as the input and produces a scalar reward as the output. The model is a 6B GPT-3 in InstructGPT, initialized by SFT with the final unembedding layer removed.

Since RM model is also a supervised model, labeled data is required for training. To this end, prompts are obtained through both the OpenAI API and manual annotation, and the SFT model generates 4 to 9 responses for each prompt. Since it is difficult to form a unified scoring standard among annotators, they tend to rank these responses to build the RM dataset, where the rankings are the labels. The dataset includes 33k training prompts. To train the RM model on this dataset, the rankings are converted to scalars since they cannot be directly used as rewards. For example, for the sorting $D > C > A = B$,

[4] https://openai.com/blog/chatgpt/

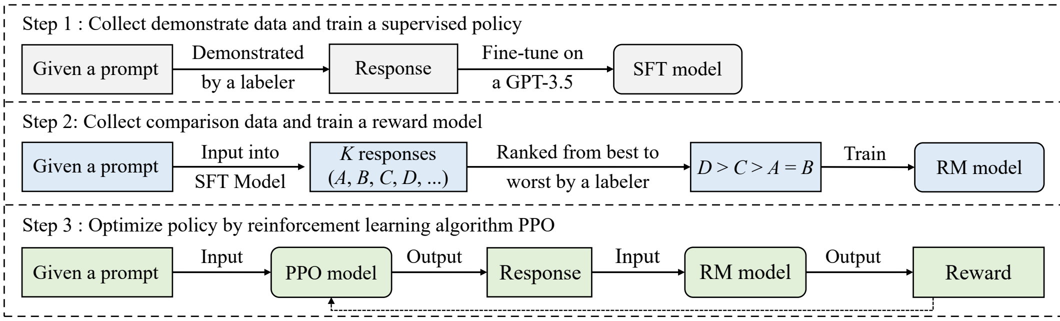


Fig. 5. The training process of InstructGPT, which is a sibling model to ChatGPT, including the SFT model, RM model, and RL model.

they assign scores of 7, 6, 4, and 4, respectively.

Formally, the loss function of the RM model is defined as follows:

$$\mathcal{L}(\theta) = -\frac{1}{\binom{K}{2}}\mathbb{E}_{(x,y_h,y_l)\sim D}[\log(\sigma(r_\theta(x,y_h) - r_\theta(x,y_l)))] \tag{3}$$

where $r_\theta(x,y)$ is a scalar produced by the RM model for the given prompt x and response y, y_h ranks higher than y_l for the prompt x, D is the RM dataset, and θ is a set of parameters in the model.

3) RL Model: Following the previous work [96], this model is fine-tuned on SFT with the proximal policy optimization (PPO) algorithm, where the input is a prompt and the output is a response. PPO serves as the agent in the RL model and is initialized with the SFT model from the first step. The environment is a prompt generator that produces a random input prompt and expects a response to the prompt. Rewards are derived from the RM model for scoring the pair of prompt and response. Formally, the RL model's objective is to maximize the following function:

$$\begin{aligned} RL(\phi) = {} & \mathbb{E}_{(x,y)\sim D_{\pi_\phi^{RL}}}[r_\theta(x,y) - \beta\log(\pi_\phi^{RL}(y|x)/\pi^{SFT}(y|x))] \\ & + \gamma\mathbb{E}_{x\sim D_{pretrain}}[\log(\pi_\phi^{RL}(x))] \end{aligned} \tag{4}$$

where π_ϕ^{RL} and π^{SFT} are the policy model and trained model, $D_{pretrain}$ is the pre-training distribution. β and γ are used to control the KL penalty and pre-training gradients. The above equation consists of three parts: a scoring function, Kullback-Leibler (KL) divergence, and a pre-training target of GPT-3. First, the RM model scores a prompt-response pair, with a higher score indicating a better response. Second, KL divergence is used to measure the distance between the distributions of responses generated by the PPO and SFT models. A smaller distance is preferable in this step since the SFT model is trained on manually labeled data, and over-optimizing it could lead to inaccurate evaluations of the responses. Finally, this part is used to calculate the probability that the RL model generates the prompt x. Note that this part contains 31k training prompts.

III. Capability Analysis of ChatGPT

By trying the ChatGPT system[5], we can easily appreciate its powerful capabilities, and at the same time, we can also find its shortcomings in various aspects. This section will conduct a tantalising glimpses capability analysis of ChatGPT based on use experiences and OpenAI's public information [17], [29].

A. Advantage Analysis and Strengths of ChatGPT

Presently, ChatGPT is initially based on GPT-3.5 and now utilizes GPT-4 of large-scale pre-trained model, and is a general-purpose human-computer conversation system obtained by combining human-labeled data and reinforcement learning training for various dialogue tasks according to the multimodal inputs of texts and images. It can understand various instructions of human beings and can complete tasks such as text generation, code writing and modification, image captioning, chart reasoning and paper summarizing etc. The specific analysis of the outstanding capabilities demonstrated by ChatGPT is as follows:

1) Multimodal (Language and Image) Understanding and Text Generation: The most intuitive feeling of using ChatGPT is that it can accurately understand the user's intention with prompts of texts (up to 2.5 million tokes) and images (still in further research), and generate various types of texts in the interactive process of dialogue. It first requires a strong multimodal understanding ability which needs to accurately understand the dialogue context and the image input. ChatGPT can not only deal with typical NLP task instructions such as classification, matching, translation, re-writing (relevant data has been processed during model training), but also deal with new tasks, which require accurate intent understanding or visual understanding, for example, organizing outputs in the form of lists and image captioning [17].

Moreover, ChatGPT has reached or exceeded human-level performance in several text generation tasks such as answering questions, providing suggestions, summarizing and polishing texts [97]. For example, when answering a question, it will not only give an accurate answer, and the generated response text will reveal the thought process involved in answering the question, and even continuously adjust and correct the answer according to the user's guidance. We believe that ChatGPT may have two reasons for this advantage: 1) the large language model trained on massive data of various forms and tasks has fully grasped language laws, and is able to understand and generate language well; 2) it has learned universal task solving ability through IFT on diverse tasks.

2) Strong Reasoning Ability and Rich Creativity: ChatGPT has good reasoning ability, especially in answering scientific questions, knowledge related questions and complex logic questions [30]. For example, ChatGPT can give the proof of a

[5] https://chat.openai.com/chat

certain theorem, and perform multi-hop reasoning according to the logical chain. ChatGPT/GPT-4 can also analyze a picture, such as ChatGPT/GPT-4 can answer the question "What can you make according to the materials in the picture?" Moreover, ChatGPT can complete various tasks in accordance with the logical chain specified by the user. Moreover, ChatGPT has rich creativity and it can generate, edit, and iteratively collaborate with users on creative and technical writing tasks such as composing music, writing scripts, or learning the user's writing style. We believe that ChatGPT may have two reasons for this advantage: 1) the model trained on code data can better perform task decomposition and logical thinking; 2) models with large-scale parameters may have formed the ability to emerge.

3) Knowledge Modeling and Planning: ChatGPT can complete almost all kinds of information consulting tasks very well, and can also achieve good results in some areas requiring professional knowledge such as medical, judicial, and financial. For example, passing the American Medical Examination [98], and the test in the Chinese environment is also good[6]. This is due to the knowledge modeling ability of ChatGPT and its basic pre-trained large model. Meanwhile, experiments have shown that GPT-4 performs at a level comparable to human performance on a variety of specialized tests and academic benchmarks. For example, it passed a simulated bar exam with a score in the top 10% of test takers, whereas GPT-3.5 scored in the bottom 10%. On this basis, ChatGPT has preliminary intelligent planning capabilities. For example, it can show the step-by-step solution ability when answering complex questions, it can show the planning ability when writing, and it can reflect the order of task execution and logical relationship when providing suggestions. It also has the ability to even provide missing details in planning goals autonomously using commonsense knowledge and reasoning [99]. We believe that knowledge modeling and acquisition is the key to the success of artificial intelligence systems. Compared with traditional knowledge engineering such as knowledge graph, which describes the limited knowledge content of a fixed schema in a symbolic way, large language model obtain model parameters and their operation processes through end-to-end self-learning, which can implicitly model richer and more comprehensive knowledge content.

In brief, ChatGPT/GPT-4 has excellent language and image understanding and generation abilities, certain reasoning and creating abilities, and powerful knowledge modeling abilities. Moreover, ChatGPT also has many other advantages and strengths, such as general task processing, machine translation for vast majority of languages, and embodied interaction, etc. To further make full use of ChatGPT, scientist can try to use it with multiple tools to solve more complex tasks for further usage.

B. Disadvantage Analysis and Limitations of ChatGPT

Although ChatGPT has powerful abilities in handling various tasks and aligning with human instructions, ChatGPT still has the following limitations.

1) Factual Errors and Hallucination Results: Although ChatGPT can integrate multiple kinds of resources to generate fluent replies, the replies generated often have factual errors (often called "hallucination problem"). Those factual errors may be due to errors and noise in the training data [100], [101]. At the same time, the internal logic and mechanism of data-driven deep learning model is still a black box for humans, so the speculation as to why the current answer was generated can neither be proven nor falsified. For some application scenarios with high accuracy requirements, such as medical consultation, the randomness of ChatGPT answers may lead to serious consequences. In addition, the ability of ChatGPT to align with human instructions comes from manually labeled data. ChatGPT does not have the ability to clarify and confirm fuzzy queries beyond the human annotations, which also brings untrustworthy obstacles to the use of ChatGPT in real scenarios. It should be noted that currently proposed technologies such as Toolformer [102] and Plugin[7] can partially alleviate the problem of factual errors, and a large amount of work has been proposed [103]–[105]. However, their effectiveness in various tasks and industries still needs to be continuously observed and explored.

2) Insufficient Modeling of Explicit Knowledge: Although a large amount of data and hundreds of billions of parameters are used, the GPT series models still only use the simplest information processing method in human wisdom, i.e., predicting the next possible word of a sentence. ChatGPT lacks the ability to produce and model accurate information. For example, there are often wrong facts and complex mathematical operations can not be performed. In fact, the advanced intelligence of human beings such as cognition and decision-making are very dependent on knowledge accumulated by social civilization for a long time. However, the GPT series models including ChatGPT have never considered the retrieval and utilization of such explicit knowledge (e.g., knowledge graphs), nor have any knowledge extracting, updating, and reusing modules.

3) Research and Development Costs are High: Achieving stable training of large models and obtaining excellent performance requires extremely high computing costs and engineering experience. OpenAI owns and utilizes the complete Microsoft Azure cloud platform to perform stable and persistent model training. For example, the training of the basic large model GPT-3 costs 12 million US dollars. In fact, the instability of large-scale data and large model training requires continuous data, code, and engineering tuning, which requires long-time technical accumulation and rich experience in system optimization.

In brief, the current ChatGPT has some limitations in terms of reliability, explicit knowledge modeling, and high research and development costs that need to be further improved. Moreover, there still exist other limitations that have not been deeply discussed in this manuscript such as data bias in political, ideological, and other area, and a lack of planning in arithmetic/reasoning questions or long text generation [106]. Scientist are making more efforts to improve these aspects.

[6] https://k.sina.com.cn/article_6208490735_1720e0cef01901d0sb.html

[7] https://openai.com/blog/chatgpt-plugins

IV. The Impact of ChatGPT Across Different Fields

ChatGPT has emerged as a popular topic of discussion across various industries [107]–[109]. As of the end of January 2023, ChatGPT boasted over 100 million monthly active users, making it the fastest-growing application in history. Hence, we next discuss the impact of ChatGPT on society from multiple aspects and analyze the key challenges for application.

From an academic perspective, ChatGPT is an important milestone in the field of AI, revealing the potential for achieving artificial general intelligence. In the past, AI research has mainly focused on the analytical capabilities of models. That is, by analyzing a given set of data, the model aims to discover the features and patterns used in practical tasks. Unlike past AI technologies, ChatGPT is a large-scale generative language model, and with the publication of GPT-4 image inputs has also been allowed, proving the feasibility of multimodal generation. ChatGPT assists humans in a range of tasks by learning and understanding human intent to generate content in a conversational manner. The emergence of ChatGPT signifies the transformation of artificial intelligence from data understanding to data generation, achieving a leap from machine perception to machine creation. Due to its powerful text generation capabilities, ChatGPT has the strong ability to produce large-scale synthetic data at a low cost. This overcomes the data limitation in the process of AI tasks and further prompts the wider application of AI technology.

From an industry perspective, ChatGPT has shown to be a valuable tool in a wide range of industries (e.g., IT, customer service, film and television, education, and medical care industries) that rely on human knowledge creation in an efficient, high-quality, and low-cost manner. With the development of generative AI technology, the model may assist or even completely replace humans to achieve most of the content creation work in the future. In essence, ChatGPT is a tool for quickly building materials. It enables low-cost or even zero-cost automated content creation, revolutionizing the content production paradigm across various industries. For instance, in the financial industry, ChatGPT can generate financial texts like investment and analysis reports to enhance work efficiency and quality. In the media industry, ChatGPT can enable intelligent writing, such as the automatic creation of news reports and articles, for creators to refine and process, significantly reducing the creative cycle. In summary, ChatGPT has a vast range of potential applications and can play a crucial role in almost any field that requires processing and understanding natural language. As technology continues to develop and innovate, ChatGPT's influence and application will continue to expand and deepen.

However, every technology is a "double-edged sword". While ChatGPT is showing great promise in related industries, it may also bring certain risks and challenges. Several examples of these challenges are as follows:

1) Intellectual Property Protection: The answer provided by ChatGPT is generated automatically, making it difficult to verify the source of the data. ChatGPT is trained on a wide range of data, including poetry, legal documents, natural conversations, blogs, and emails. These data inevitably contain copyrighted information. Consequently, ChatGPT's responses may over-reference other people's work or articles, potentially leading to infringement disputes.

2) Safety Aspects: ChatGPT is easily to be used to generate misleading information or phishing emails for cyber scams at scale. At the same time, it can also help criminals find security holes in websites and generate network attack scripts faster and easier. In other words, generative AI will undoubtedly greatly reduce the threshold of cyber attacks, because it not only expands the number of potential threats, but also empowers novices to participate in security attacks.

3) Ethics and Integrity: Due to ChatGPT's high efficiency and high quality of response, it surpasses most of the existing problem-solving software. However, its diverse responses to the same question make it difficult to detect plagiarism or cheating. This could lead to students and researchers using the tool to cheat, which could have adverse consequences for teaching and academic integrity. Furthermore, abusing ChapGPT may cause more social ethics problems (e.g., generating harmful contents) without proper mitigation measures.

4) Environmental Impact: Since ChatGPT involves a huge amount of parameters and pre-training data, it consumes a significant amount of hardware resources during training. Providing ChatGPT services to millions of users every day also generates carbon emissions, which accumulate daily and are challenging to estimate.

Hence, to develop generative AI technology responsibly, safely, and controllably, one should prioritize the following concepts for achieving high-quality, healthy, and sustainable development. First, transparency and accountability are crucial. The application and development of ChatGPT need to be open and transparent to ensure that society and relevant stakeholders understand its use and potential risks. Moreover, one should establish accountability mechanisms to ensure that ChatGPT users and developers are responsible for its use and application. Second, one should follow moral and legal principles. The development and use of ChatGPT must align with ethical and legal principles. One needs to ensure that its application does not violate ethical and legal regulations, such as those related to fraud, invasion of privacy, discrimination, and other undesirable purposes. Third, one needs to focus on environmental protection and sustainability. The use of ChatGPT can consume significant computing resources and energy, leading to substantial carbon emissions. Therefore, one should pay attention to environmental protection and sustainability, reduce the impact on the environment as much as possible, and apply the technology in a more sustainable way.

V. The Potential Future Development Trends of ChatGPT

Even GPT-4 brings significant improvements for ChatGPT, several core problems are also still not been solved yet. Thus, this section analyzes the future development of ChatGPT. First we discuss the hallucination problem, existing in most of the large-scale models, including ChatGPT. Then, considering the huge cost and the large-scale hardware requirements, we discuss how to reduce the size of language models for model compression. Finally, we propose several other trends in future, and the overview diagram is shown in Fig. 6.

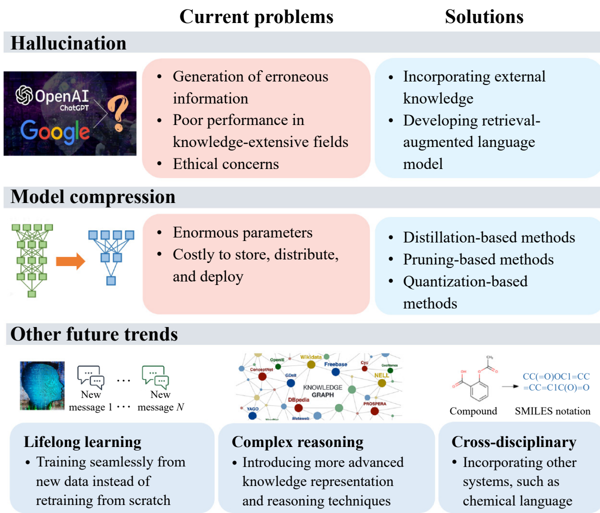


Fig. 6. The future trends of ChatGPT for hallucination, model compression and other important trends. Image materials are adapted from the Internet[8].

A. Solving the Hallucination Problem of ChatGPT

Recent advances in large-scale language models (LLM), have enabled the generation of seemingly high-quality language from appropriately formulated prompts. However, these models are known to generate "hallucinated" facts, which can lead to misleading conclusions. The potential for generating erroneous information remains a significant concern in the use of such models, which limits their applicability in knowledge-extensive fields such as finance, legal advisement and medical suggestions. In response to this problem, some researchers have proposed incorporating external knowledge sources to mitigate "hallucination" and the "retrieval-augmented language" model has emerged as a popular approach [110]–[112]. Typically, these systems retrieve relevant knowledge from a large knowledge base such as Wikipedia or other web texts, given a prompt, and generate a response by leveraging the retrieved knowledge. By incorporating external knowledge sources, such as named entities or specific facts, these systems can improve their accuracy and reduce the potential for generating misleading or incorrect information.

However, the use of external knowledge sources also poses challenges, such as the need for effective retrieval methods and the possibility of introducing bias into the models. Therefore, further research is needed to improve the effectiveness of these approaches and to address potential ethical concerns when deploying large-scale language models in various applications.

B. Model Compression for LLM

Researchers have shown that the capabilities of transformer-based language models grow log-linearly with the number of parameters [113], [114], and some promising abilities such as in-context learning [25] and chain-of-thoughts [57] emerge when the model size exceeds certain thresholds. Equipped with the scaling law [114], today's modern large language models have revolutionized natural language processing with their remarkable performance on various language tasks. However, these models come with a significant cost. They usually contain over 100 billion parameters, which presents challenges for practical usage, including increased costs for storage, distribution, and deployment in real-world applications. As a result, researchers need to develop new approaches for model compression and optimization to make these models more practical and accessible for real-world use cases.

Reducing the size of language models while maintaining their performance on downstream tasks has been a long-standing challenge in the field, and researchers have proposed multiple approaches to tackle this problem. Distillation-based approaches, such as those proposed in [49], [115]–[117], involve training a smaller student model using extra training data or soft labels generated by a larger teacher model (LLM or ChatGPT). Teacher model by hundreds or even thousands of times while maintaining the emerging properties remains an open problem that requires further research. Pruning-based approaches [118], [119] aim to reduce the size of the model by removing a number of unimportant weights, while achieving similar performance to the original model. Quantization-based approaches [120], [121] use 8-bit or even binary numbers to store the model weights, thereby reducing the model size compared to 32-bit storage. Although these approaches successfully reduce the model size, special hardware design is often required to accelerate model inference speed, which can limit their usage. Please refer to [122] for a more comprehensive survey of the field.

C. Other Future Trends

ChatGPT has a vast potential for potential future development from both research and practical perspectives, and it is impossible to list them all in this paper. As discussed in the previous section, we have provided a few examples to illustrate the wide range of possibilities for the model's future direction.

1) Lifelong Learning Incorporation: Experiments have proved that ChatGPT published in January 2023 can solve 70% theory of mind (ToM) tasks, which is comparable with that of seven-year-old children [123]. However, this does mean that ChatGPT has mind as humans and it is just the results of strong capability of the language models. Message of current affairs may forever not be learned by ChapGPT if we stop updating the model. To adapt to such changes, one possible approach is to incorporate lifelong learning [124] into the ChatGPT model. By doing so, the model can seamlessly incorporate new data without requiring retraining from scratch. This is particularly advantageous for large-scale language models such as ChatGPT, as it saves computing resources and enables more efficient and effective performance. Additionally, lifelong learning can enable the model to evolve in real-time, leading to more personalized and relevant results. This is particularly beneficial in ChatGPT, where user interactions are constantly evolving, and up-to-date informa-

[8] https://www.quora.com/Will-Chat-GPT-take-over-Google, https://www.appnovation.com/blog/seo-google-algorithm-and-knowledge-graph

tion is critical. Overall, incorporating lifelong learning into ChatGPT can lead to improved performance and more accurate results, benefiting both researchers and users alike.

2) Complex Reasoning Abilities Enhancement: While ChatGPT already has impressive reasoning abilities, there is always room for improvement to tackle even more complex problems. Further research can focus on enhancing the reasoning capabilities of the model by incorporating advanced knowledge representation and reasoning techniques, such as knowledge graphs [125], [126] and logical reasoning [127]. These improvements help the model better understand and reason over complex structures and relationships between different concepts. By advancing the reasoning abilities of ChatGPT, the model can become even more powerful and versatile in its applications.

3) Cross-Disciplinary Integration: The significant language processing performance of ChapGPT also has the potential to be fused with chemical system. Simplified molecular input linear entry system (SMILES) [128] is composed of specific ASCII symbols which is used to define chemical, representing the compositions and structure for molecules. SMILES can also be seen as a kind of language system, and thus it is also deserved to be chained with ChatGPT for further research.

In addition to the above points, we have also identified several other future trends or directions of ChatGPT: 1) realizing a larger long-term memory may help to better understand and process more complex tasks, such as reading a book; 2) improving the robustness and sensitivity to inputs is important since the prompts and their sequence may greatly influence the outputs, leading to suboptimal or non-aligned results; 3) improving transparency, interpretability and consistency is also important, as it helps to establish better trust or collaboration with the user; and 4) external calls should be considered for application expansion. There is still much work to be done to create a better ChatGPT model or an AGI model.

VI. Conclusion

In this paper, we provide a brief overview on the recently released AI agent ChatGPT, including its predecessor, strengths and the limitations, social impact and potential future development. We discuss the core techniques of ChatGPT, mainly including large-scale language models, in-context learning, reinforcement learning from human feedback and their potential relationship. In a nutshell, ChatGPT is a phenomenal technology product that holds significant importance in both academic and industry domains.

However, it is still unclear how ChatGPT can realize such powerful functions by combing simple algorithmic components of gradient descent, large-scale language models constructed by Transformer, and vast amounts of data. One important future direction is to research the phenomenon of emergence in large language models to understand how the emergence appears and what kind of abilities does it may have. This could be important in creating stronger AIGC models.

Currently, more and more companies and research groups are following OpenAI's lead in developing their own ChatGPT-like products or AIGC products. For example, Microsoft has combined ChatGPT with their search engine Bing to improve the quality of the search results, Baidu has published their ChatGPT aliked robot called ERNIE Bot, which can generate images based on text descriptions, and Sensetime has developed their SenseChat robot, which can generate figures, videos, and 3D contents. ChatGPT-related technologies have attracted worldwide attention and it becomes a major force in computer science.

In brief, the essence of ChatGPT-like products is still a kind of AIGC instead of reaching artificial super intelligence (ASI). In this way, people should further focus the side-effects of technology products, such as the reliability and validity of the answers by artificial intelligence and the potential for cheating to occur. Overall, when people are looking forward to more powerful artificial intelligence, consensus on ethical and responsible usage are also required to be carefully considered.

References

[1] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Q. Zhou, W. Li, and J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 1, p. 140, Jan. 2020.

[2] H. Y. Du, Z. H. Li, D. Niyato, J. W. Kang, Z. H. Xiong, X. M. Shen, and D. I. Kim, "Enabling AI-generated content (AIGC) services in wireless edge networks," arXiv preprint arXiv: 2301.03220, 2023.

[3] Y. Ming, N. N. Hu, C. X. Fan, F. Feng, J. W. Zhou, and H. Yu, "Visuals to text: A comprehensive review on automatic image captioning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1339–1365, Aug. 2022.

[4] C. Q. Zhao, Q. Y. Sun, C. Z. Zhang, Y. Tang, and F. Qian, "Monocular depth estimation based on deep learning: An overview," *Sci. China Technol. Sci.*, vol. 63, no. 9, pp. 1612–1627, Jun. 2020.

[5] J. Lü, G. H. Wen, R. Q. Lu, Y. Wang, and S. M. Zhang, "Networked knowledge and complex networks: An engineering view," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1366–1383, Aug. 2022.

[6] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[7] J. Chen, K. L. Wu, Y. Yu, and L. B. Luo, "CDP-GAN: Near-infrared and visible image fusion via color distribution preserved GAN," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 9, pp. 1698–1701, Sept. 2022.

[8] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proc. 38th Int. Conf. Machine Learning*, 2021, pp. 8748–8763.

[9] L. Yang, Z. L. Zhang, Y. Song, S. D. Hong, R. S. Xu, Y. Zhao, W. T. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," arXiv preprint arXiv: 2209.00796, 2022.

[10] C. C. Leng, H. Zhang, G. R. Cai, Z. Chen, and A. Basu, "Total variation constrained non-negative matrix factorization for medical image registration," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 5, pp. 1025–1037, May 2021.

[11] S. Reed, K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-Maron, M. Gimenez, Y. Sulsky, J. Kay, J. T. Springenberg, T. Eccles, J. Bruce, A. Razavi, A. Edwards, N. Heess, Y. T. Chen, R. Hadsell, O. Vinyals, M. Bordbar, and N. de Freitas, "A generalist agent," arXiv preprint arXiv: 2205.06175, 2022.

[12] Y. Liu, Y. Shi, F. H. Mu, J. Cheng, and X. Chen, "Glioma segmentation-oriented multi-modal MR image fusion with adversarial learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1528–1531, Aug. 2022.

[13] J. Gusak, D. Cherniuk, A. Shilova, A. Katrutsa, D. Bershatsky, X. Y. Zhao, L. Eyraud-Dubois, O. Shlyazhko, D. Dimitrov, I. Oseledets, and

O. Beaumont, "Survey on large scale neural network training," arXiv preprint arXiv: 2202.10435, 2022.

[14] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," arXiv preprint arXiv: 2204.06125, 2022.

[15] U. Singer, A. Polyak, T. Hayes, X. Yin, J. An, S. Y. Zhang, Q. Y. Hu, H. Yang, O. Ashual, O. Gafni, D. Parikh, S. Gupta, and Y. Taigman, "Make-a-video: Text-to-video generation without text-video data," arXiv preprint arXiv: 2209.14792, 2022.

[16] W. X. Jiao, W. X. Wang, J.-T. Huang, X. Wang, and Z. P. Tu, "Is ChatGPT a good translator? Yes with GPT-4 as the engine," arXiv preprint arXiv: 2301.08745, 2023.

[17] OpenAI, "Gpt-4 technical report," 2023. [Online]. Available: https://cdn.openai.com/papers/gpt-4.pdf.

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

[19] M. Khosla, A. Anand, and V. Setty, "A comprehensive comparison of unsupervised network representation learning methods," arXiv preprint arXiv: 1903.07902, 2019.

[20] Q. Y. Sun, C. Q. Zhao, Y. Ju, and F. Qian, "A survey on unsupervised domain adaptation in computer vision tasks," *Sci. Sinica Technol.*, vol. 52, no. 1, pp. 26–54, 2022.

[21] C. Ieracitano, A. Paviglianiti, M. Campolo, A. Hussain, E. Pasero, and F. C. Morabito, "A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 1, pp. 64–76, Jan. 2021.

[22] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, pp. 9, 2019.

[23] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.

[24] Y. Q. Wang, Q. M. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, p. 63, May 2021.

[25] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 1877–1901.

[26] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Machine Learning*, Sydney, Australia, 2017, pp. 1126–1135.

[27] J. Beck, R. Vuorio, E. Z. Liu, Z. Xiong, L. Zintgraf, C. Finn, and S. Whiteson, "A survey of meta-reinforcement learning," arXiv preprint arXiv: 2301.08028, 2023.

[28] Q. X. Dong, L. Li, D. M. Dai, C. Zheng, Z. Y. Wu, B. B. Chang, X. Sun, J. J. Xu, L. Li, and Z. F. Sui, "A survey on in-context learning," arXiv preprint arXiv: 2301.00234, 2022.

[29] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, "Training language models to follow instructions with human feedback," arXiv preprint arXiv: 2203.02155, 2022.

[30] C. W. Qin, A. Zhang, Z. S. Zhang, J. A. Chen, M. Yasunaga, and D. Y. Yang, "Is ChatGPT a general-purpose natural language processing task solver?" arXiv preprint arXiv: 2302.06476, 2023.

[31] C. Stokel-Walker and R. Van Noorden, "What ChatGPT and generative AI mean for science," *Nature*, vol. 614, no. 7947, pp. 214–216, Feb. 2023.

[32] C. Stokel-Walker, "ChatGPT listed as author on research papers: Many scientists disapprove," *Nature*, vol. 613, no. 7945, pp. 620–621, Jan. 2023.

[33] C. X. Zhai, *Statistical Language Models for Information Retrieval*. Hanover, USA: Now Publishers Inc., 2008, pp. 1–141.

[34] Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," in *Proc. 13th Int. Conf. Neural Information Processing Systems*, Denver, USA, 2000, pp. 893–899.

[35] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Nov. 2011.

[36] Y. M. Ju, Y. Z. Zhang, K. Liu, and J. Zhao, "Generating hierarchical explanations on text classification without connecting rules," arXiv preprint arXiv: 2210.13270, 2022.

[37] M. J. Zhu, Y. X. Weng, S. Z. He, K. Liu, and J. Zhao, "Learning to answer complex visual questions from multi-view analysis," in *Proc. 7th China Conf. Knowledge Graph and Semantic Computing*, Qinhuangdao, China, 2022, pp. 154–162.

[38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. 1st Int. Conf. Learning Representations*, Scottsdale, USA, 2013.

[39] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, USA, 2013, pp. 3111–3119.

[40] T. Mikolov, M. Karafiat, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. 11th Annu. Conf. Int. Speech Communication Association*, Makuhari, Japan, 2010, pp. 1045–1048.

[41] M. Sundermeyer, R. Schlueter, and H. Ney, "LSTM neural networks for language modeling," in *Proc. 13th Annu. Conf. Int. Speech Communication Association*, Portland, USA, 2012, pp. 194–197.

[42] X. Qiu, T. X. Sun, Y. G. Xu, Y. F. Shao, N. Dai, and X. J. Huang, "Pre-trained models for natural language processing: A survey," *Sci. China Technol. Sci.*, vol. 63, no. 10, pp. 1872–1897, Sept. 2020.

[43] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proc. Conf. North American Chapter of the Association for Computational Linguistics*, New Orleans, USA, 2018, pp. 2227–2237.

[44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, USA, 2019, pp. 4171–4186.

[45] M. Lewis, Y. H. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880.

[46] Y. H. Liu, M. Ott, N. Goyal, J. F. Du, M. Joshi, D. Q. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv: 1907.11692, 2019.

[47] Z. Z. Lan, M. D. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," in *Proc. 8th Int. Conf. Learning Representations*, Addis Ababa, Ethiopia, 2020.

[48] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv preprint arXiv: 1910.01108, 2019.

[49] X. Q. Jiao, Y. C. Yin, L. F. Shang, X. Jiang, X. Chen, L. L. Li, F. Wang, and Q. Liu, "TinyBERT: Distilling BERT for natural language understanding," in *Proc. Findings of the Association for Computational Linguistics*, 2020, pp. 4163–4174.

[50] Y. M. Cui, W. X. Che, T. Liu, B. Qin, and Z. Q. Yang, "Pre-training with whole word masking for Chinese BERT," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3504–3514, Nov. 2021.

[51] W. J. Liu, P. Zhou, Z. Zhao, Z. R. Wang, Q. Ju, H. T. Deng, and P. Wang, "K-BERT: Enabling language representation with knowledge graph," in *Proc. 34th AAAI Conf. Artificial Intelligence*, New York, USA, 2019, pp. 2901–2908.

[52] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer in BERTology: What we know about how BERT works," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, 2020.

[53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, USA, 2017, pp. 6000–6010.

[54] W. Q. Ren, Y. Tang, Q. Y. Sun, C. Q. Zhao, and Q.-L. Han, "Visual semantic segmentation based on few/zero-shot learning: An overview," arXiv preprint arXiv: 2211.08352, 2022.

[55] S. Smith, M. Patwary, B. Norick, P. LeGresley, S. Rajbhandari, J. Casper, Z. Liu, S. Prabhumoye, G. Zerveas, V. Korthikanti, E. Zhang, R. Child, R. Y. Aminabadi, J. Bernauer, X. Song, M. Shoeybi, Y. X. He, M. Houston, S. Tiwary, and B. Catanzaro, "Using DeepSpeed and megatron to train megatron-turing NLG 530B, a large-scale generative language model," arXiv preprint arXiv: 2201.11990, 2022.

[56] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. Von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, L. Fei-Fei, C. Finn, T. Gale, L. Gillespie, K. Goel, N. Goodman, S. Grossman, N. Guha, T. Hashimoto, P. Henderson, J. Hewitt, D. E. Ho, J. Hong, K. Hsu, J. Huang, T. Icard, S. Jain, D. Jurafsky, P. Kalluri, S. Karamcheti, G. Keeling, F. Khani, O. Khattab, P. W. Koh, M. Krass, R. Krishna, R. Kuditipudi, A. Kumar, F. Ladhak, M. Lee, T. Lee, J. Leskovec, I. Levent, X. L. Li, X. C. Li, T. Y. Ma, A. Malik, C. D. Manning, S. Mirchandani, E. Mitchell, Z. Munyikwa, S. Nair, A. Narayan, D. Narayanan, B. Newman, A. Nie, J. C. Niebles, H. Nilforoshan, J. Nyarko, G. Ogut, L. Orr, I. Papadimitriou, J. S. Park, C. Piech, E. Portelance, C. Potts, A. Raghunathan, R. Reich, H. Y. Ren, F. Rong, Y. Roohani, C. Ruiz, J. Ryan, C. Ré, D. Sadigh, S. Sagawa, K. Santhanam, A. Shih, K. Srinivasan, A. Tamkin, R. Taori, A. W. Thomas, F. Tramèr, R. E. Wang, W. Wang, B. H. Wu, J. J. Wu, Y. H. Wu, S. M. Xie, M. Yasunaga, J. X. You, M. Zaharia, M. Zhang, T. Y. Zhang, X. K. Zhang, Y. H. Zhang, L. Zheng, K. Zhou, and P. Liang, "On the opportunities and risks of foundation models," arXiv preprint arXiv: 2108.07258, 2021.

[57] J. Wei, X. Z. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," arXiv preprint arXiv: 2201.11903, 2022.

[58] J. Huang and K. C.-C. Chang, "Towards reasoning in large language models: A survey," arXiv preprint arXiv: 2212.10403, 2022.

[59] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," arXiv preprint arXiv: 2205.11916, 2022.

[60] Z. S. Zhang, A. Zhang, M. Li, and A. Smola, "Automatic chain of thought prompting in large language models," arXiv preprint arXiv: 2210.03493, 2022.

[61] D. Zhou, N. Scharli, L. Hou, J. Wei, N. Scales, X. Z. Wang, D. Schuurmans, C. Cui, O. Bousquet, Q. Le, and E. Chi, "Least-to-most prompting enables complex reasoning in large language models," arXiv preprint arXiv: 2205.10625, 2022.

[62] X. Z. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," arXiv preprint arXiv: 2203.11171, 2022.

[63] Y. X. Weng, M. J. Zhu, F. Xia, B. Li, S. Z. He, K. Liu, and J. Zhao, "Large language models are reasoners with self-verification," arXiv preprint arXiv: 2212.09561, 2022.

[64] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, Y. X. Li, X. Z. Wang, M. Dehghani, S. Brahma, A. Webson, S. S. Gu, Z. Y. Dai, M. Suzgun, X. Y. Chen, A. Chowdhery, A. Castro-Ros, M. Pellat, K. Robinson, D. Valter, S. Narang, G. Mishra, A. Yu, V. Zhao, Y. P. Huang, A. Dai, H. K. Yu, S. Petrov, E. H. Chi, J. Dean, J. Devlin, A. Roberts, D. Zhou, Q. V. Le, and J. Wei, "Scaling instruction-finetuned language models," arXiv preprint arXiv: 2210.11416, 2022.

[65] D. Goldwasser and D. Roth, "Learning from natural instructions," *Mach. Learn.*, vol. 94, no. 2, pp. 205–232, Feb. 2014.

[66] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," in *Proc. 10th Int. Conf. Learning Representations*, 2022.

[67] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press, 2018.

[68] L. Xue, C. Y. Sun, D. Wunsch, Y. J. Zhou, and F. Yu, "An adaptive strategy via reinforcement learning for the prisoner's dilemma game," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 301–310, Jan. 2018.

[69] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018.

[70] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Y. Wang, T. Pfaff, Y. H. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, Oct. 2019.

[71] Y. B. Jin, X. W. Liu, Y. C. Shao, H. T. Wang, and W. Yang, "High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning," *Nat. Mach. Intell.*, vol. 4, no. 12, pp. 1198–1208, Dec. 2022.

[72] A. Ecoffet, J. Huizinga, J. Lehman, K. O. Stanley, and J. Clune, "First return, then explore," *Nature*, vol. 590, no. 7847, pp. 580–586, Feb. 2021.

[73] R. F. Wu, Z. K. Yao, J. Si, and H. H. Huang, "Robotic knee tracking control to mimic the intact human knee profile based on actor-critic reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 1, pp. 19–30, Jan. 2021.

[74] S. K. Gottipati, B. Sattarov, S. F. Niu, Y. Pathak, H. R. Wei, S. C. Liu, K. J. Thomas, S. Blackburn, C. W. Coley, J. Tang, S. Chandar, S. Chandar, and Y. Bengio, "Learning to navigate the synthetically accessible chemical space using reinforcement learning," in *Proc. 37th Int. Conf. Machine Learning*, 2020, pp. 344.

[75] J. K. Wang, C.-Y. Hsieh, M. Y. Wang, X. R. Wang, Z. X. Wu, D. J. Jiang, B. B. Liao, X. J. Zhang, B. Yang, Q. J. He, D. S. Cao, and T. J. Hou, "Multi-constraint molecular generation based on conditional transformer, knowledge distillation and reinforcement learning," *Nat. Mach. Intell.*, vol. 3, no. 10, pp. 914–922, Oct. 2021.

[76] Y. N. Wan, J. H. Qin, X. H. Yu, T. Yang, and Y. Kang, "Price-based residential demand response management in smart grids: A reinforcement learning-based approach," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 1, pp. 123–134, Jan. 2022.

[77] W. Z. Liu, L. Dong, D. Niu, and C. Y. Sun, "Efficient exploration for multi-agent reinforcement learning via transferable successor features," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 9, pp. 1673–1686, Sept. 2022.

[78] J. Y. Weng, H. Y. Chen, D. Yan, K. C. You, A. Duburcq, M. H. Zhang, Y. Su, H. Su, and J. Zhu, "Tianshou: A highly modularized deep reinforcement learning library," *J. Mach. Learn. Res.*, vol. 23, no. 267, pp. 1–6, Aug. 2022.

[79] J. Schulman, S. Levine, P. Moritz, M. Jordan, and P. Abbeel, "Trust region policy optimization," in *Proc. 32nd Int. Conf. Int. Conf. Machine Learning*, Lille, France, 2015, pp. 1889–1897.

[80] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," arXiv preprint arXiv: 1812.05905, 2018.

[81] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv: 1707.06347, 2017.

[82] S. Fujimoto, H. Van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. 35th Int. Conf. Machine Learning*, Stockholm, Sweden, 2018, pp. 1582–1591.

[83] X. Y. Chen, C. Wang, Z. J. Zhou, and K. W. Ross, "Randomized ensembled double q-learning: Learning fast without a model," in *Proc. 9th Int. Conf. Learning Representations*, 2021.

[84] P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, USA, 2017, pp. 4302–4310.

[85] W. B. Knox and P. Stone, "TAMER: Training an agent manually via evaluative reinforcement," in *Proc. 7th IEEE Int. Conf. Development and Learning*, Monterey, USA, 2008, pp. 292–297.

[86] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *Proc. 34th Int. Conf. Machine Learning*, Sydney, Australia, 2017, pp. 2285–2294.

[87] G. Warnell, N. R. Waytowich, V. Lawhern, and P. Stone, "Deep TAMER: Interactive agent shaping in high-dimensional state spaces," in *Proc. 32nd AAAI Conf. Artificial Intelligence*, New Orleans, USA, 2018, pp. 1545–1554.

[88] A. Glaese, N. McAleese, M. Trębacz, J. Aslanides, V. Firoiu, T. Ewalds, M. Rauh, L. Weidinger, M. Chadwick, P. Thacker, L. Campbell-Gillingham, J. Uesato, P.-S. Huang, R. Comanescu, F. Yang, A. See, S. Dathathri, R. Greig, C. Chen, D. Fritz, J. S. Elias, R.

Green, S. Mokrá, N. Fernando, B. X. Wu, R. Foley, S. Young, I. Gabriel, W. Isaac, J. Mellor, D. Hassabis, K. Kavukcuoglu, L. A. Hendricks, and G. Irving, "Improving alignment of dialogue agents via targeted human judgements," arXiv preprint arXiv: 2209.14375, 2022.

[89] D. Cohen, M. Ryu, Y. Chow, O. Keller, I. Greenberg, A. Hassidim, M. Fink, Y. Matias, I. Szpektor, C. Boutilier, and G. Elidan, "Dynamic planning in open-ended dialogue using reinforcement learning," arXiv preprint arXiv: 2208.02294, 2022.

[90] J. Kreutzer, S. Khadivi, E. Matusov, and S. Riezler, "Can neural machine translation be improved with user feedback?" in *Proc. Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, USA, 2018, pp. 92–105.

[91] S. Kiegeland and J. Kreutzer, "Revisiting the weaknesses of reinforcement learning for neural machine translation," in *Proc. Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 1673–1681.

[92] W. C. S. Zhou and K. Xu, "Learning to compare for better training and evaluation of open domain natural language generation models," in *Proc. 34th AAAI Conf. Artificial Intelligence*, New York, USA, 2020, pp. 9717–9724.

[93] E. Perez, S. Karamcheti, R. Fergus, J. Weston, D. Kiela, and K. Cho, "Finding generalizable evidence by learning to convince Q&A models," in *Proc. Conf. Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. Natural Language Processing*, Hong Kong, China, 2019, pp. 2402–2411.

[94] A. Madaan, N. Tandon, P. Clark, and Y. M. Yang, "Memory-assisted prompt editing to improve GPT-3 after deployment," in *Proc. Conf. Empirical Methods in Natural Language Processing*, Abu Dhabi, United Arab Emirates, 2022, pp. 2833–2861.

[95] C. Lawrence and S. Riezler, "Improving a neural semantic parser by counterfactual learning from human bandit feedback," in *Proc. 56th Annu. Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 2018, pp. 1820–1830.

[96] N. Stiennon, L. Ouyang, J. Wu, D. M. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. F. Christiano, "Learning to summarize with human feedback," in *Proc. 34th Neural Information Processing Systems*, 2020, pp. 3008–3021.

[97] P. Liang, R. Bommasani, T. Lee, D. Tsipras, D. Soylu, M. Yasunaga, Y. A. Zhang, D. Narayanan, Y. H. Wu, A. Kumar, B. Newman, B. H. Yuan, B. Yan, C. Zhang, C. Cosgrove, C. D. Manning, C. Ré, D. Acosta-Navas, D. A. Hudson, E. Zelikman, E. Durmus, F. Ladhak, F. Rong, H. Y. Ren, H. X. Yao, J. Wang, K. Santhanam, L. Orr, L. Zheng, M. Yuksekgonul, M. Suzgun, N. Kim, N. Guha, N. Chatterji, O. Khattab, P. Henderson, Q. Huang, R. Chi, S. M. Xie, S. Santurkar, S. Ganguli, T. Hashimoto, T. Icard, T. Y. Zhang, V. Chaudhary, W. Wang, X. C. Li, Y. F. Mai, Y. H. Zhang, and Y. Koreeda, "Holistic evaluation of language models," arXiv preprint arXiv: 2211.09110, 2022.

[98] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, and V. Tseng, "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," *PLoS Digit. Health*, vol. 2, no. 2, p. e0000198, 2023.

[99] Y. Q. Xie, C. Yu, T. Y. Zhu, J. B. Bai, Z. Gong, and H. Soh, "Translating natural language to planning goals with large-language models," arXiv preprint arXiv: 2302.05128, 2023.

[100] A. Borji, "A categorical archive of ChatGPT failures," arXiv preprint arXiv: 2302.03494, 2023.

[101] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, "Mathematical capabilities of ChatGPT," arXiv preprint arXiv: 2301.13867, 2023.

[102] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom, "Toolformer: Language models can teach themselves to use tools," arXiv preprint arXiv: 2302.04761, 2023.

[103] W. X. Zhou, S. Zhang, H. Poon, and M. Chen, "Context-faithful prompting for large language models," arXiv preprint arXiv: 2303.11315, 2023.

[104] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Y. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. M. Yang, S. Welleck, B. P. Majumder, S. Gupta, A. Yazdanbakhsh, and P. Clark, "Self-refine: Iterative refinement with self-feedback," arXiv preprint arXiv: 2303.17651, 2023.

[105] B. Paranjape, S. Lundberg, S. Singh, H. Hajishirzi, L. Zettlemoyer, and M. T. Ribeiro, "ART: Automatic multi-step reasoning and tool-use for large language models," arXiv preprint arXiv: 2303.09014, 2023.

[106] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Z. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, "Sparks of artificial general intelligence: Early experiments with GPT-4," arXiv preprint arXiv: 2303.12712, 2023.

[107] F.-Y. Wang, J. Yang, X. X. Wang, J. J. Li, and Q.-L. Han, "Chat with ChatGPT on industry 5.0: Learning and decision-making for intelligent industries," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 4, pp. 831–834, Apr. 2023.

[108] F.-Y. Wang, Q. H. Miao, X. Li, X. X. Wang, and Y. L. Lin, "What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 575–579, Mar. 2023.

[109] Q. H. Miao, W. B. Zheng, Y. S. Lv, M. Huang, W. W. Ding, and F.-Y. Wang, "DAO to HANOI via DeSci: AI paradigm shifts from AlphaGo to ChatGPT," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 4, pp. 877–897, Apr. 2023.

[110] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, "Retrieval augmented language model pre-training," in *Proc. 37th Int. Conf. Machine Learning*, 2020, pp. 3929–3938.

[111] P. S. H. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. 34th Advances in Neural Information Processing Systems*, 2020, pp. 9459–9474.

[112] Y. Z. Zhang, S. Q. Sun, X. Gao, Y. W. Fang, C. Brockett, M. Galley, J. F. Gao, and B. Dolan, "RetGen: A joint framework for retrieval and grounded text generation modeling," in *Proc. 36th AAAI Conf. Artificial Intelligence*, 2022, pp. 11739–11747.

[113] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, "Emergent abilities of large language models," arXiv preprint arXiv: 2206.07682, 2022.

[114] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," arXiv preprint arXiv: 2001.08361, 2020.

[115] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv: 1503.02531, 2015.

[116] S. Q. Sun, Y. Cheng, Z. Gan, and J. J. Liu, "Patient knowledge distillation for BERT model compression," in *Proc. Conf. Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. Natural Language Processing*, Hong Kong, China, 2019, pp. 4323–4332.

[117] Z. Q. Sun, H. K. Yu, X. D. Song, R. J. Liu, Y. M. Yang, and D. Zhou, "MobileBERT: A compact task-agnostic BERT for resource-limited devices," in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics*, 2020, pp. 2158–2170.

[118] M. Gordon, K. Duh, and N. Andrews, "Compressing BERT: Studying the effects of weight pruning on transfer learning," in *Proc. 5th Workshop on Representation Learning for NLP*, 2020, pp. 143–155.

[119] T. L. Chen, J. Frankle, S. Y. Chang, S. J. Liu, Y. Zhang, Z. Y. Wang, and M. Carbin, "The lottery ticket hypothesis for pre-trained BERT networks," in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 1328.

[120] S. Shen, Z. Dong, J. Y. Ye, L. J. Ma, Z. W. Yao, A. Gholami, M. W. Mahoney, and K. Keutzer, "Q-BERT: Hessian based ultra low precision quantization of BERT," in *Proc. 34th AAAI Conf. Artificial Intelligence*, New York, USA, 2020, pp. 8815–8821.

[121] H. L. Bai, W. Zhang, L. Hou, L. F. Shang, J. Jin, X. Jiang, Q. Liu, M. Lyu, and I. King, "BinaryBERT: Pushing the limit of BERT quantization," in *Proc. 59th Annu. Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. Natural Language Processing*, 2020, pp. 4334–4348.

[122] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," arXiv preprint arXiv: 1710.09282, 2017.

[123] M. Kosinski, "Theory of mind may have spontaneously emerged in large language models," arXiv preprint arXiv: 2302.02083, 2023.

[124] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural*

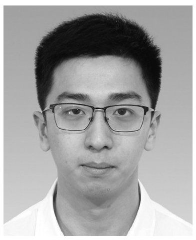
Netw., vol. 113, pp. 54–71, May 2019.

[125] S. X. Ji, S. R. Pan, E. Cambria, Marttinen, and S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2022.

[126] Y. Y. Lan, S. Z. He, K. Liu, and J. Zhao, "Knowledge reasoning via jointly modeling knowledge graphs and soft rules," arXiv preprint arXiv: 2301.02781, 2023.

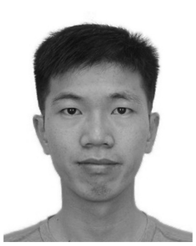
[127] Y. J. Bang, S. Cahyawijaya, N. Lee, W. L. Dai, D. Su, B. Wilie, H. Lovenia, Z. W. Ji, T. Z. Yu, W. Chung, Q. V. Do, Y. Xu, and P. Fung, "A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity," arXiv preprint arXiv: 2302.04023, 2023.

[128] D. Weininger, "SMILEs, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988.



Tianyu Wu received the B.S. degree in automation from East China University of Science and Technology in 2020, where he is currently pursuing the Ph.D. degree with control science and engineering. His fields of interests include molecular representation learning, reinforcement learning, and natural language processing.

Shizhu He received the Ph.D. degree in pattern recognition and intelligent system from Institute of Automation, Chinese Academy of Sciences, in 2016. Currently, he is an Associate Researcher in Institute of Automation, Chinese Academy of Sciences. His research interests focus on question answering, deep learning, and natural language processing.



Jingping Liu received the Ph.D. degree from the School of Computer Science, Fudan University, in 2021. He now is a Lecturer at the School of Information Science and Engineering, East China University of Science and Technology. His research interests include pre-trained language model, knowledge graph, and data mining.

Siqi Sun received the Ph.D. degree in computer science at the Toyota Technological Institute in Chicago, USA. He now holds a position as a Young Investigator at the Research Institute of Intelligent Complex Systems at Fudan University. His primary research interests include AI for science and natural language processing.

Kang Liu received the Ph.D. degree in pattern recognition and intelligent system from Institute of Automation, Chinese Academy of Sciences, in 2010. He is currently a Professor in Institute of Automation, Chinese Academy of Sciences. His research interests include knowledge graph, deep learning, and natural language processing.

Qing-Long Han (Fellow, IEEE) received the B.Sc. degree in mathematics from Shandong Normal University in 1983, and the M.Sc. and Ph.D. degrees in control engineering from East China University of Science and Technology, in 1992 and 1997, respectively.

Professor Han is Pro Vice-Chancellor (Research Quality) and a Distinguished Professor at Swinburne University of Technology, Australia. He held various academic and management positions at Griffith University and Central Queensland University, Australia. His research interests include networked control systems, multi-agent systems, time-delay systems, smart grids, unmanned surface vehicles, and neural networks.

Professor Han was awarded The 2021 Norbert Wiener Award (the Highest Award in systems science and engineering, and cybernetics) and The 2021 M. A. Sargent Medal (the Highest Award of the Electrical College Board of Engineers Australia). He was the recipient of The 2022 IEEE SMC Society Andrew P. Sage Best Transactions Paper Award, The 2021 IEEE/CAA Journal of Automatica Sinica Norbert Wiener Review Award, The 2020 IEEE Systems, Man, and Cybernetics Society Andrew P. Sage Best Transactions Paper Award, The 2020 IEEE Transactions on Industrial Informatics Outstanding Paper Award, and The 2019 IEEE Systems, Man, and Cybernetics Society Society Andrew P. Sage Best Transactions Paper Award.

Professor Han is a Member of the Academia Europaea (The Academy of Europe). He is a Fellow of The International Federation of Automatic Control (IFAC), a Fellow of The Institution of Engineers Australia (IEAust) and a Fellow of Chinese Association of Automation (CAA). He is a Highly Cited Researcher in both Engineering and Computer Science (Clarivate). He has served as an AdCom Member of IEEE Industrial Electronics Society (IES), a Member of IEEE IES Fellows Committee, and Chair of IEEE IES Technical Committee on Networked Control.

Yang Tang (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Donghua University, in 2006 and 2010, respectively. From 2008 to 2010, he was a Research Associate with The Hong Kong Polytechnic University, Hong Kong, China. From 2011 to 2015, he was a Post-Doctoral Researcher with the Humboldt University of Berlin, Germany, and with the Potsdam Institute for Climate Impact Research, Germany. He is now a Professor with the East China University of Science and Technology. His current research interests include distributed estimation/control/optimization, cyber-physical systems, hybrid dynamical systems, computer vision, reinforcement learning and their applications.

Prof. Tang was a recipient of the Alexander von Humboldt Fellowship and has been the ISI Highly Cited Researchers Award by Clarivate Analytics from 2017. He is a Senior Board Member of Scientific Reports, an Associate Editor of *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Industrial Informatics*, *IEEE Transactions on Circuits and Systems-I: Regular Papers*, *IEEE Transactions on Cognitive and Developmental Systems*, *IEEE Transactions on Emerging Topics in Computational Intelligence*, *IEEE Systems Journal*, *Engineering Applications of Artificial Intelligence* (IFAC Journal) and *Science China Information Sciences*, etc. He is a (leading) Guest Editor for several special issues focusing on autonomous systems, robotics, and industrial intelligence in *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Cognitive and Developmental Systems*, *IEEE Transactions on Emerging Topics in Computational Intelligence*.