SCIENCE CHINA
Information Sciences

• PERSPECTIVE •

June 2023, Vol. 66 166201:1–166201:3
https://doi.org/10.1007/s11432-023-3740-x

Feedback is all you need: from ChatGPT to autonomous driving

Hong CHEN[1,2,3*], Kang YUAN[1], Yanjun HUANG[3,4], Lulu GUO[1], Yulei WANG[1] & Jie CHEN[2,3*]

[1]*College of Electronic and Information Engineering, Tongji University, Shanghai* 201804*, China;*
[2]*National Key Laboratory of Autonomous Intelligent Unmanned Systems, Shanghai* 201210*, China;*
[3]*Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai* 201210*, China;*
[4]*School of Automotive Studies, Tongji University, Shanghai* 201804*, China*

Received 21 March 2023/Accepted 30 March 2023/Published online 23 April 2023

Citation Chen H, Yuan K, Huang Y J, et al. Feedback is all you need: from ChatGPT to autonomous driving. Sci China Inf Sci, 2023, 66(6): 166201, https://doi.org/10.1007/s11432-023-3740-x

The birth of chat generative pre-trained transformer (ChatGPT). In November 2022, Microsoft and OpenAI jointly released an artificial intelligence chatbot program called ChatGPT [1]. In just two months, the number of global users surpassed 100 million, making it the fastest-growing consumer application in history. ChatGPT can answer almost any question like an encyclopedia, with an amazing level of fluency, a human-like expression, and even sophistication. Surprisingly, it can generate complete and coherent texts, write code like a junior engineer, and even pass medical and legal licensing exams, demonstrating the emergence of machine intelligence.

Technologies behind ChatGPT. First, ChatGPT chooses an ideal application: conversation, which could occur anywhere and anytime for anyone. ChatGPT enables people to easily experience the allure of artificial intelligence (AI) through natural question-and-answer interactions, the creative composition of textual content, and a powerful knowledge base, sparking a surge of interest in AI's potential. Indeed, related technologies have already demonstrated breakthroughs and feasibility. "G" stands for generative models, which can randomly generate observable data and can be used to model different types of data such as images, text, and sound. As the typical generative model, autoregressive language models (ALMs) are widely used to generate text automatically. "P" stands for pre-training, which is the process of obtaining a trained model from massive labeled data that may be independent of specific tasks. Since Google released bidirectional encoder representations from transformers (BERT) in 2018, "pre-training + fine-tuning" has gradually gained traction in the natural language processing (NLP) field. The letter "T" stands for transformer, a deep learning architecture that can incorporate attention mechanisms and perform exceptionally well when processing long text. *Furthermore, ChatGPT employs the technology of reinforcement learning with human feedback (RLHF).* It employs human experts to rank and label the answers generated by the pre-trained language model with prompt inputs. These labels are then used to create a comparison database for training a reward model of reinforcement learning (RL) and to optimize the language model. In this way, the model output can be aligned with human expressions, logic, and common sense. These technologies are precisely why ChatGPT's performance has progressed from quantitative change to qualitative change, culminating in the emergence of intelligence.

Recall of feedback: always the most efficient mechanism to deal with an open, complex, varying, and uncertain environment. As early as the Qin Dynasty in China, the design of Dujiangyan reflected the concept of feedback: the Minjiang River is divided into an inner and outer stream through the "Yuzui Dike", with the former being deep and narrow and the latter being shallow and wide. This constructs a feedback mechanism to control the water into the Chengdu Plain: the majority of the water flows to the outer stream during the wet season, protecting the Chengdu Plain from flooding. During the dry season, the water flows into the inner stream and irrigates the Chengdu Plain. In 1948, Norbert Wiener published the famous "Cybernetics: Or Control and Communication in the Animal and the Machine". He investigated the information flow in artificial and living systems, clarified the control mechanisms of these systems in general, and finally summarized feedback into basic principle [2]. *Feedback is defined as the process of returning the output of a system, comparing it with reference, and correcting the input in some way that affects the system's output.* Conventionally, as depicted in Figure 1, industrial systems measure physical variables, such as temperature, pressure, flow, force, and acceleration, as feedback to pursue imitation and replacement of human operations. It is the feedback that deals with plant uncertainties/disturbances and guarantees the great

* Corresponding author (email: chenhong2019@tongji.edu.cn, chenjie206@tongji.edu.cn)

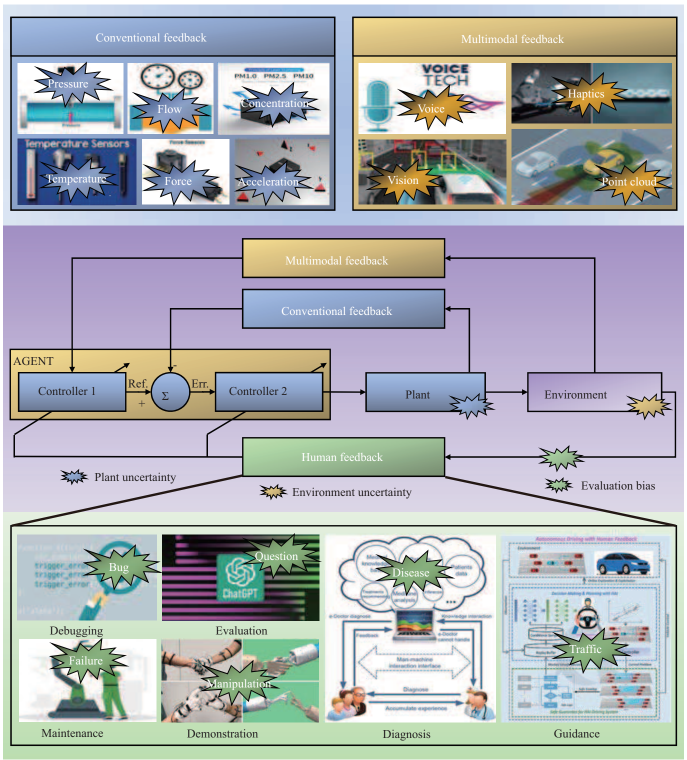

Figure 1 Feedback mechanism and its applications.

success of industrial control. Nowadays, intelligent systems use multimodal sensing as feedback to mimic, enhance and replace human perception, cognition, communication, creation, and other abilities. In an autonomous driving (AD) system, for example, signals from the camera, radar, Lidar, etc. are introduced in the feedback channel, and in either hierarchical or end-to-end frameworks to realize the way of feedback affecting the system's output. In the hierarchical perception-planning-control framework, the planning module (Controller 1 in Figure 1) receives multimodal feedback from the camera, radar, Lidar, etc. to detect traffic states. It then makes a decision and plans a desired trajectory to a tracker (Controller 2) which finally outputs the control signal to vehicle actuators based on the feedback from the vehicle's speed, acceleration, and position sensors. While in the end-to-end framework, as is the case with most vision-driven models, images are fed back to the AD agent to directly control the vehicle. It is noted that in a transportation system, the agent could be pedestrians, vehicles with and without human-driver, and other traffic participants. Other types of signals such as text, voice, and haptics are also widely used as multimodal feedback. Consequently, feedback has broadened from single-modal signals to multimodal signals, protecting the system from environmental uncertainties.

Extension of feedback: human feedback, further expands the denotation of feedback. As the system and the interacting environment become more and more complex, it is not sufficient for controllers/agents to rely on multimodal sensing to function properly and safely. As the most intelligent species on the planet, humans can perform well in complex and uncertain environments. Consequently, as in the lower part of Figure 1, a human is introduced into the feedback loop to augment system intelligence in terms of controller design or assisting the AI agent's learning process via human calibration, manipulation, debugging, supervision, knowledge, evaluation, and cognition, to deal with both plant and environment uncertainties, as well as evaluation biases. Indeed, the concept of human feedback has long permeated various fields such as automation, computer, and artificial intelligence [3]. For example, in industrial control systems, sensors require annual inspection and calibration by human experts to address issues such as bias and aging. The widely applied PID controller is tuned by human engineers via evaluating system outputs. Computer programs necessitate extensive manual debugging and testing by humans. While in the field of AD, subjective assessment methods are commonly used by automotive engineers to analyze the actual performance and reliability of AD systems to adjust the parameters of the vehicle controller. Furthermore, AI-based electronic doctor systems learn from doctors' feedback on diagnostic results of various diseases to constantly improve their diagnostic capabilities, accuracy, and efficiency. Image processing employs feedback from humans to fine-tune the deep neural network (DNN) model for enhancing prediction accuracy. Then, returning to ChatGPT, while its success is superficially due to a surge in parameters and large models, as ChatGPT's predecessor, GPT-3's performance remains unstable. Regardless of how parameters of GPT-3 are increased, the improvement in model performance is very limited, making it extremely difficult to solve the long-tail effect of training even with "brute force computations". To resolve this problem, ChatGPT uses RLHF technology, making it an engineering innovation based on previous technological breakthroughs, with the innovation being the incorporation of human feedback. It provides an intelligence-augmenting paradigm to deal with the openness and uncertainty of dialog scenarios. *Therefore, in the presence of GPT technologies, it is exactly human feedback that becomes the key to ChatGPT's success.*

The ways human feedback is injected into the sys-

tem include two types: online and offline. We may call the way used in ChatGPT an offline one, where reliable human feedback is collected for training the reward model. On the other hand, human feedback can also be introduced online to adjust controllers or train agents, similar to conventional feedback in industrial control systems. In both cases, as illustrated in Figure 1, it must be the top priority to evaluate the quality of the feedback data. For example, in robot systems, learning from reliable interactions with humans improves the efficiency and performance of robots. In uncertain chat scenarios, the human may provide biased or even incorrect answers, resulting in an unreliable alignment standard. Here, online human feedback could degrade the machine's performance. Thus, ChatGPT is precisely driven by reward functions learned from human experts' subjective feelings and reliable evaluations gathered offline to make a splash. As a result, ensuring the reliability of human feedback is the most important issue arising from using feedback mechanisms in learning-based methods.

Another success of human feedback: autonomous driving. Autonomous vehicles are expected to outperform human drivers in terms of reducing fatal accidents. However, the current state of the art in AD is still far from what is expected. Both rule-based and model-based technologies, which are induced by scenarios, as well as AI-based methods, are insufficiently intelligent to deal with open, dynamic, and infinite scenarios in the real world, particularly for safety-critical AD applications [4]. Taking reinforcement learning for example, it is driven by a reward function and learns how to drive a car by interacting with traffic environment [5]. Because of the dynamically changing and complex driving environment, however, sampling is inefficient, learning is expensive, and designing rewards is difficult. In a similar vein to ChatGPT, given the strong robustness and adaptability of human drivers in complex driving scenarios, humans are also expected to have significant potential in improving the performance of AI-based methods in real-world open traffic. Indeed, we present two of our studies on using human feedback to augment the RL driving ability through online training. Furthermore, the proposed methods can be applied based on pre-trained policies using offline human driving data.

To improve the learning speed and personalized ability of decision-making and control in AD while achieving safe and continuous evolution, we propose a novel framework by developing a hybrid augmented intelligence (HAI) method to incorporate human feedback into the learning process. A decision-making scheme is developed based on interactive reinforcement learning (Int-RL). And a human driver assesses the learning level of the RL agent in real time. When the learning is unreasonable or the learning speed is slow, human guidance is used to replace the machine's decisions and assist the learning process. These decisions will then be passed to the planning layer to perform the longitudinal and lateral motion planning tasks using model predictive control, respectively. Finally, a safety guarantee mechanism is proposed to ensure the safety of the HAI system, which includes establishing a safe driving envelope and designing a safe exploration/exploitation logic based on trial-and-error on the desired decision. Results show that it can realize an efficient, reliable, and safe evolution of AD to pursue a higher traveling efficiency.

Furthermore, to improve the data efficiency of the RL method for AD in complex scenarios and enhance the ability of vehicles to actively conduct cut-ins in congested traffic scenarios, we propose a human-knowledge-augmented mandatory lane change method. Specifically, in typical congested off-ramp scenarios, we first developed an exploration mechanism with a safety guarantee. By encoding the prior knowledge of the human driver and constraining the effective range of the state space, safety is ensured. Then, a human expert provides online feedback to the learning system of AD. The unreasonable behavior of the vehicle can thus be discovered in time by the human and guided by taking control. Based on the goal of the driving task and human guidance, a reward/policy enhancement mechanism is built to help the vehicle accelerate policy learning. Finally, the decision-making instruction is sent to the downstream control module to complete the vehicle's control task. According to the experimental results, the proposed method can improve the data efficiency, the training speed, and the success rate of active cut-in under various traffic densities.

In the field of AD, safe and reliable human feedback can not only enhance the vehicle's learning efficiency but also shape a safe, human-like, and trustworthy AD system. Conversely, it may drastically degrade the performance of AD, even causing functional failure and triggering safety accidents. Qualified and trustworthy evaluators can lessen the adverse effects of this problem. To further learn driving skills from ordinary drivers or even passengers, designing an online feedback mechanism by evaluating the reliability of human feedback is a critical issue to address. Aside from technological challenges, legal issues should also be focused on. Due to the introduction of human factors, legal issues of responsibility allocation inevitably emerge when traffic accidents occur. Quantifying or qualifying the positive and negative impacts of human factors on accidents or even establishing new legal regulations must be seriously considered in the process of widely implementing human-centered intelligent systems.

In conclusion, conventional physical variables (single-modal sensing) and multimodal sensing are used as feedback in a closed-loop manner to improve system function and performance. Gradually, human feedback is introduced to align agents with human beings under more complex and uncertain environments. It is anticipated that, in the age of big data, large models, and powerful computing, the feedback mechanism will pave the way for AI to achieve artificial general intelligence. Feedback is all you need!

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2022YFB2502900) and Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100).

References

1 Thorp H H. ChatGPT is fun, but not an author. Science, 2023, 379: 313

2 Wiener N. Cybernetics or Control and Communication in the Animal and the Machine. Cambridge: MIT Press, 2019

3 Zheng N, Liu Z, Ren P, et al. Hybrid-augmented intelligence: collaboration and cognition. Front Inf Technol Electron Eng, 2017, 18: 153–179

4 Feng S, Sun H, Yan X, et al. Dense reinforcement learning for safety validation of autonomous vehicles. Nature, 2023, 615: 620–627

5 Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: a survey. IEEE Trans Intell Transp Syst, 2021, 23: 4909–4926