EAI Endorsed Transactions
on AI and Robotics

Research Article EAI.EU

# An Overview of OpenAI's Sora and Its Potential for Physics Engine Free Games and Virtual Reality

Zuyan Chen [1], Shuai Li[2,3,*], Md. Asraful Haque[4]

[1]Department of Electronic and Electrical Engineering, Swansea University, Fabian Way, Swansea, SA1 8EN, UK
[2]Faculty of Information Technology and Electrical Engineering, University of Oulu, Pentti Kaiteran katu 1, Oulu, 90570, Finland
[3]VTT-Technical Research Centre of Finland, VTT, Kaitoväylä 1, Oulu, 90570, Finland
[4]Computational Unit, Z.H. College of Engg. & Technology, Aligarh Muslim University, Aligarh-202002, India

## Abstract

Sora, OpenAI's latest text-to-video model, is particularly skilled at understanding the physical world, and all of the content it generates mostly consistent with the laws of physics. This indicates that Sora already has the beginnings of a world model and has the potential to become an excellent physics engine in the near future. This paper analyses and explains in detail the potential applications of Sora in physics engines and virtual reality. In addition, its advantages and disadvantages over traditional physics engines are compared based on its unique behavioural characteristics. Finally, it looks forward to the application of Sora in other fields.

Received on 01 March 2024; accepted on 05 March 2024; published on 06 March 2024

**Keywords:** Sora, World Model, GPT, Physics Engine, Virtual Reality, Game, Simulator, Text To Video, Video Generation.



doi:10.4108/airo.5273

## 1. Introduction

In recent years, generative models, especially large language models (LLM) and diffusion models, have wildly succeeded in their domains. Large language models, such as GPT, LLAMA, and PALM, have been applied to text generation as well as a wide range of industries [1–3]. Surprisingly, language seems to be the carrier of human knowledge, and the capabilities of big language models are diffused into programming, clerical, design, engineering, and other fields. Human interaction with the models has become more accessible, and the ability to utilize big models has increased human productivity [4, 5]. The diffusion model has given rise to the flourishing development of the field of text-generated diagrams, and when human beings marvel at the delicacy of its generative effect, they suddenly realize that it seems to understand the laws of existence of the physical world and the generated content is surprisingly in line with the objective laws, allowing it to be utilized in simulating the world [6, 7].

*Corresponding author. Email: shuai.li@oulu.fi

Given its profound grasp of language, visual perception, and physical dynamics, Sora represents a major breakthrough in artificial intelligence and video production [8]. It demonstrates how AI may be used to provide immersive and interesting material for a range of markets, including communication, education, entertainment, and the arts. "Sora" means "sky" in Japanese. The question is whether it will provide users with unlimited creative opportunities or not.

### 1.1. Sora

Sora is the latest release of OpenAI's text to video model, and as you can see from the details published in the technical report, it's certainly another huge breakthrough in the process of AI's understanding of the world [8]. The results show that Sora can generate high-quality videos of up to 60 seconds from the text provided by user. Meanwhile Sora is capable of extending to generate images or videos based on uploaded images or videos. From the generation samples presented, it can be found that Sora is particularly good at understanding the physical world, and all the generated content mostly compliant with the laws of physics. There is every indication that Sora has the beginnings of a world model and the potential to become an excellent physics engine in the future [9].

### 1.2. Physics Engine

A physics engine is a type of software used to simulate a variety of physical properties, often used in fields that require the use of dynamics, acoustics, optics, and electromagnetism [10]. This includes object collisions, particle systems, optical rendering, and many other physical behaviors. However, the more accurate the physics engine, the more computation it requires, so it is always necessary to choose the appropriate physics engine for different application scenarios to achieve better performance [11].

### 1.3. Video Game

Video games are an important scenario for the application of physics engines [12]. Unlike the 2D games of the past, 3D games tend to give players an immersive experience. Since the release of the first 3D game in the late 20th century, physics engine iterations have often caused an uproar. This is because, in order to support the real-time nature of the game as well as the feel of the experience, physics engines have had to choose a balance between performance and effectiveness. However, with the rapid development of computer hardware, especially after the introduction of GPUs, the effectiveness of physics engines has been qualitatively improved, and the gaming industry has entered a period of booming development [13].

### 1.4. Virtual Reality

Virtual Reality is a virtual 3D space created by a computer that allows the user to interact with any object in that space [14]. The physics engine plays an extremely important role here; it seems to be the eternal law of virtual reality, controlling the physical behavior of the virtual world. Any user interaction data and the current state of the space are fed into the physics engine in real time, and then rendered in real time according to the physics engine's response [15].

## 2. The Sora Based World Model

As mentioned above, Sora has the potential to be a world model. The difference between physical interaction based on a world model and traditional interaction based on a physics engine is indicated in Fig 1 [16]. The traditional physics engine computational process requires the prior construction of a physical space, and obtains the state of the object as well as the current action to perform a large number of numerical simulations, and finally obtains an approximation of the corresponding, and then returns the dynamics results to the object. The world model-based physics engine, on the other hand, does not need to construct a space in advance, but only needs the current state and object's action to generate the subsequent physical state of the object. The extreme freedom of the world model makes it possible to build an open world [17]. As OpenAI demonstrated, the Minecraft game video generated by Sora already indicates the initial characteristics of the physics engine, and in the future it may be possible to approximate the effects of the physics engine via utilizing prompt engineering to allow Sora to generate videos that are more focused on the physics properties.

## 3. Advantages

1. **Generalization**: The Sora-based physics engine is more generative, and once the model is trained, it can adapt to different physical conditions without the need to update the physics computation rules for it, and do not require to build the physics scenario in advance.

2. **Effectiveness**: World models are able to capture more subtle physical changes due to the fact that traditional physics rule-based engines have explicit mathematical and physical theorems, which makes some complicate interactions extremely difficult to simulate. World models, by contrast, are driven by large amounts of data and can learn minute characteristics.

3. **Innovation**: The world model can also be utilized with simulations of phenomena that are currently physically difficult to explain, and often these behaviours cannot be reproduced with conventional physics engines. Simulators built in this way can make simulations more robust.

## 4. Weakness

1. **Explainability**: Due to the problem of Explainability of the world model, this leads to certain times when the results produced by the model are not controllable, e.g. generating phenomena that do not exist objectively.

2. **Data Quality**: Large models often require numerous data for training, but how to ensure the quality of the data is a critical issue; good data enables superior model performance, however bad data makes the model uncontrollable.

3. **Real-time**: In general, a world model is a model with a large amount of parameters, which requires vast computations, causing real-time a very essential issue. Ensuring temporal coherence in video

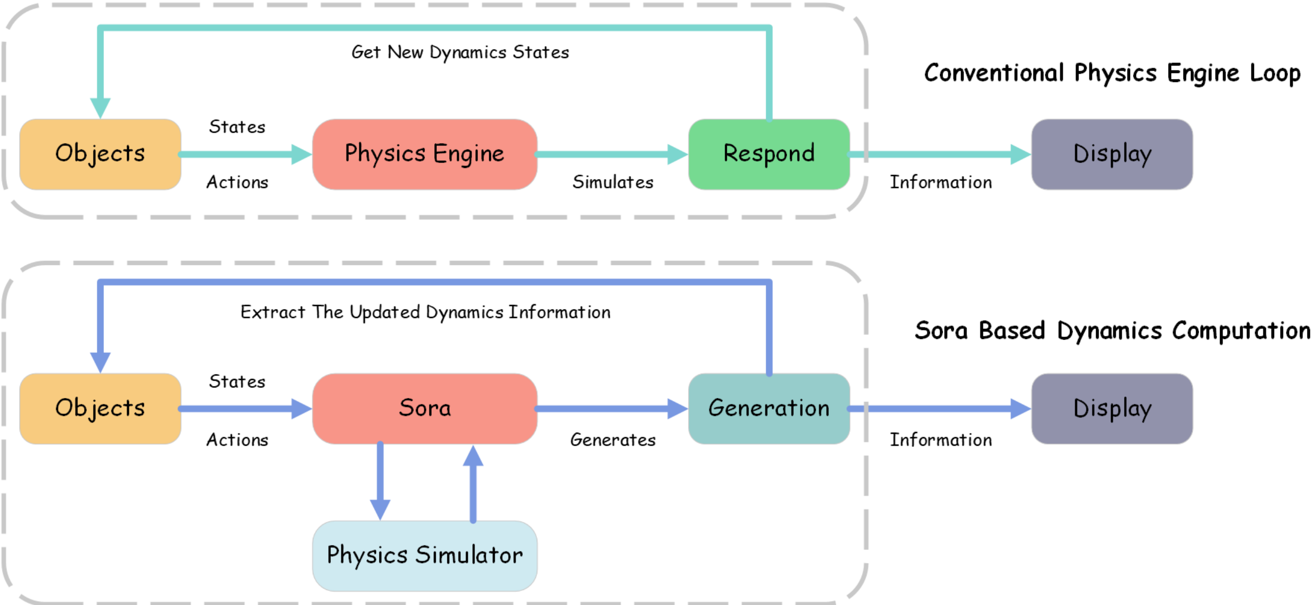


**Figure 1.** The Comparison Between Conventional Physics Engine Loop And Sora Based Dynamics Computation.

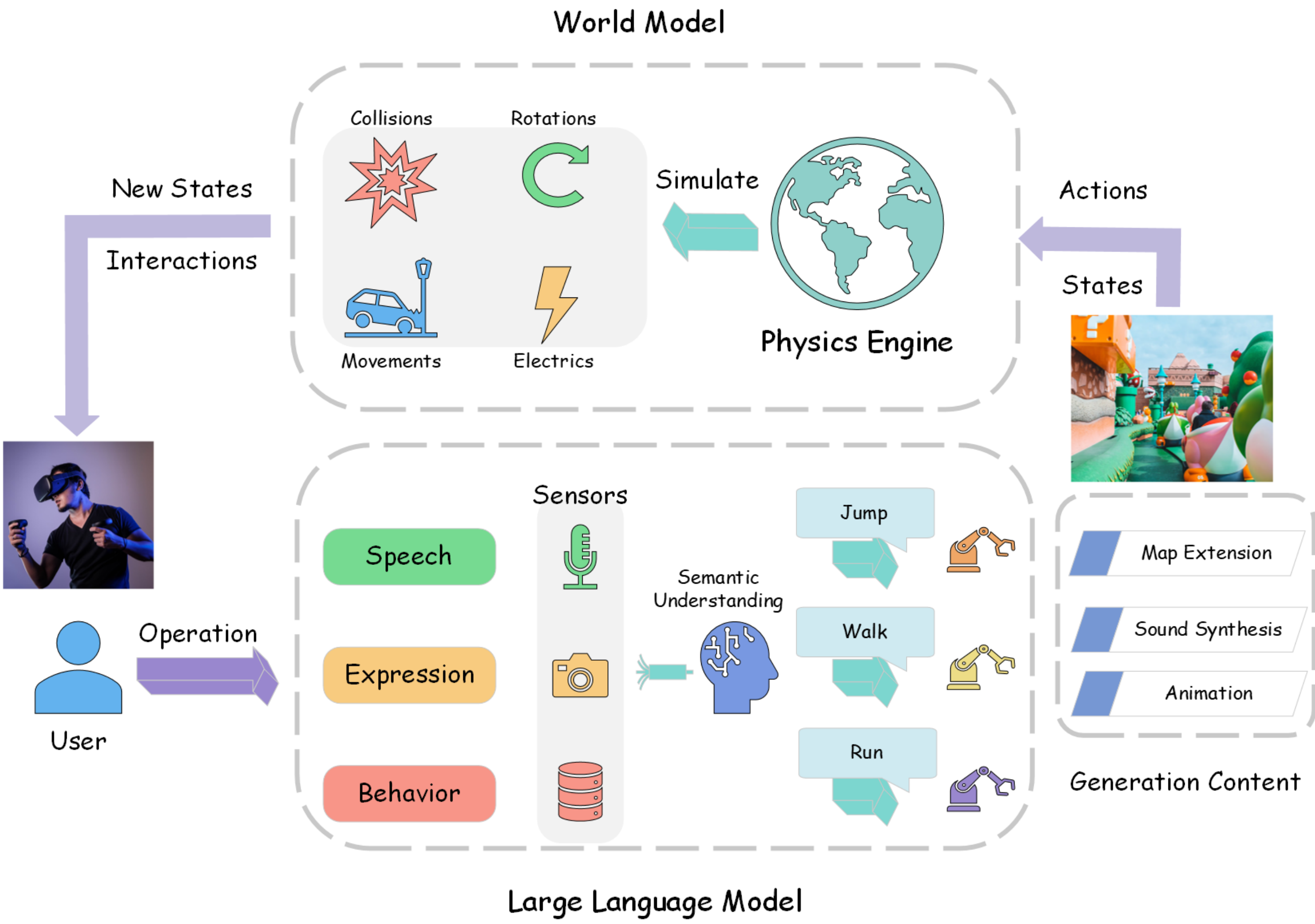


**Figure 2.** The Potential Structure For Future Immersive Interaction.

generation is crucial for producing seamless and natural-looking sequences. Maintaining consistency in actions, movements, and transitions over time poses a challenge for text-to-video converters, especially for long and complex narratives.

## 5. Future Outlook

Overall, the emergence of Sora opens a new door for physics engines and virtual reality, allowing for a more general model of the world. In future applications, Sora can be combined with other generative models, such as the LLM, to build a fully immersive virtual world, the structure of which is shown in Fig. 2. The user interacts with the LLM via a series of sensors, which extracts relevant semantic information and translates it into different behaviours in the virtual world. Through these specific behaviours, pre-information can be generated for input to the world model, such as map generation, sound synthesis, animation display, etc. The world model receives this information, performs simulation calculations, and finally generates new state and interaction data and feeds it back to the user side. In addition, world models can be used in robot simulation, autonomous driving, animation and many other fields. We expect to witness the qualitative changes that Sora will bring to various industries in the near future [18, 19].

## References


[1] Floridi, L. and Chiriatti, M. (2020) Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines* **30**: 681–694.

[2] Anil, R., Dai, A.M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., Shakeri, S. *et al.* (2023) Palm 2 technical report. *arXiv preprint arXiv:2305.10403* .

[3] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y. and Narasimhan, K. (2024) Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems* **36**.

[4] Haque, M.A. and Li, S. (2023) The potential use of chatgpt for debugging and bug fixing. *EAI Endorsed Transactions on AI and Robotics* **2**(1): e4–e4.

[5] Haque, M.A. and Li, S. (2024) Exploring chatgpt and its impact on society. *AI and Ethics* : 1–13.

[6] Dhariwal, P. and Nichol, A. (2021) Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **34**: 8780–8794.

[7] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022) High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: 10684–10695.

[8] Brooks, T., Peebles, B., Homes, C., DePue, W., Guo, Y., Jing, L., Schnurr, D. *et al.* (2024) Video generation models as world simulators URL `https://openai.com/research/video-generation-models-as-world-simulators`.

[9] Ha, D. and Schmidhuber, J. (2018) World models. *arXiv preprint arXiv:1803.10122* .

[10] Todorov, E., Erez, T. and Tassa, Y. (2012) Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems* (IEEE): 5026–5033.

[11] Degrave, J., Hermans, M., Dambre, J. *et al.* (2019) A differentiable physics engine for deep learning in robotics. *Frontiers in neurorobotics* : 6.

[12] Haas, J.K. (2014) A history of the unity game engine. *Diss. Worcester Polytechnic Institute* **483**(2014): 484.

[13] Qiu, W. and Yuille, A. (2016) Unrealcv: Connecting computer vision to unreal engine. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14* (Springer): 909–916.

[14] Hummel, J., Wolff, R., Stein, T., Gerndt, A. and Kuhlen, T. (2012) An evaluation of open source physics engines for use in virtual reality assembly simulations. In *Advances in Visual Computing: 8th International Symposium, ISVC 2012, Rethymnon, Crete, Greece, July 16-18, 2012, Revised Selected Papers, Part II 8* (Springer): 346–357.

[15] Maciel, A., Halic, T., Lu, Z., Nedel, L.P. and De, S. (2009) Using the physx engine for physics-based virtual surgery with force feedback. *The International Journal of Medical Robotics and Computer Assisted Surgery* **5**(3): 341–353.

[16] Hafner, D., Lillicrap, T., Norouzi, M. and Ba, J. (2020) Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193* .

[17] Peng, X.B., Andrychowicz, M., Zaremba, W. and Abbeel, P. (2018) Sim-to-real transfer of robotic control with dynamics randomization. In *2018 IEEE international conference on robotics and automation (ICRA)* (IEEE): 3803–3810.

[18] Wu, J., Yildirim, I., Lim, J.J., Freeman, B. and Tenenbaum, J. (2015) Galileo: Perceiving physical object properties by integrating a physics engine with deep learning. *Advances in neural information processing systems* **28**.

[19] Koenig, N. and Howard, A. (2004) Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566)* (IEEE), **3**: 2149–2154.