Sora for Smart Mining: Towards Sustainability With Imaginative Intelligence and Parallel Intelligence

Yuting Xie, Cong Wang, Kunhua Liu, Zhe Xuanyuan, Yuhang He, Hui Cheng, Andreas Nüchter, Lingxi Li, Rouxing Huai, Shuming Tang, Siji Ma, Long Chen, *Senior member, IEEE*

***Abstract*—This letter summarizes discussions from IEEE TIV's Autonomous Mining Workshop, emphasizing the potential of video generation models in advancing smart mining.**

***Index Terms*—Smart Mining, Video Generation Models, Sustainable Mining, Intelligent Vehicles.**

This letter is resulted from IEEE TIV's Autonomous Mining Workshop. We have successfully conducted over 10 seminars on autonomous mining, and this summary highlights key components discussed in recent seminars regarding how technologies like Sora promote autonomous mining.

I. Looking into Sora: can it advance smart mining?

Sora [1] has recently garnered attention for its ability to generate 60-second videos depicting complex scenes. Over the past year, research on similar models has progressed rapidly, showing significant improvements in visual fidelity and coherence. Despite limitations, these advancements inspire confidence in the technology's future. This letter explores potential applications and challenges in autonomous mining stemming from these developments.

A. Potential Opportunities

Mining Scene Understanding. Mining environments are inherently complex, marked by irregularity, low lighting, and cluttered objects, presenting challenges for scene understanding. Video generation models have shown proficiency in producing coherent visual signals, suggesting their grasp of spatial and temporal relationships among objects. Delving deeper into these scene cognition abilities observed in video generation models could improve visual signal processing in intricate settings like mining.

Human Machine Interaction. Text-to-video models bridge the gap between text and video, crafting videos from text and summarizing videos into text. In smart mining, when enriched with mining knowledge, these models can generate animated demonstrations of mining strategies, enhancing visualization. They could also analyze mining operation videos in real-time, assisting engineers in decision-making and thereby enhancing safety while reducing workload.

Scenario Engineering. Scenario engineering, a cutting-edge theoretical framework for refining AI, integrates model learning and testing across real and virtual scenarios. Video generation models, grounded in the mining domain, can be employed to produce customized mining videos serving as virtual scenarios within this framework, enhancing the reliability of autonomous mining.

End-To-End Autonomous Mining. Given the recent advancements in video generation models, it has been suggested that these models demonstrate a profound understanding of global dynamics, potentially making simulation engines obsolete. Leveraging these video generation models for world modeling shows great potential for visual end-to-end autonomous mining [2].

B. Limitations & Challenges

Hallucinations. The issue of hallucinations is prevalent among large foundation models, referring to the generation of meaningless or deviating content from the provided source material. Currently, finding a complete solution for this problem remains elusive. Addressing how to validate AI inferences and mitigate these hallucinations in autonomous mining applications requires meticulous consideration.

Physical Paradoxes. Sora and similar models face challenges in simulating physical laws and understanding cause-and-effect relationships. For example, they may struggle to

(Corresponding author: Long Chen.)

Yuting Xie and Hui Cheng are with the School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: xieyt8@mail2.sysu.edu.cn; chengh9@mail.sysu.edu.cn).

Cong Wang is with the School of Automotive Studies, Tongji University, Shanghai 201804, China (e-mail: 1953104@tongji.edu.cn).

Kunhua Liu is with the School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao 266520, China. (e-mail: liukunhua@qut.edu.cn).

Zhe Xuanyuan is with the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai 519087, China (e-mail: zhexuanyuan@uic.edu.cn).

Yuhang He is with the Department of Computer Science, University of Oxford, Oxford OX1 2JD, UK (e-mail: yuhang.he@cs.ox.ac.uk).

Andreas Nüchter is with Computer Science (Robotics), University of Würzburg, Germany (e-mail: andreas.nuechter@uni-wuerzburg.de).

Lingxi Li is with Indiana University-Purdue University, Indianapolis IN 46202, USA (e-mail: ll7@iupui.edu).

Rouxing Huai is with Beijing Huairou Academy of Parallel Sensing, Beijing 101499, China (e-mail: xr.huai@qaii.ac.cn).

Shuming Tang is affiliated with the Institute of Automation, Chinese Academy of Sciences in Beijing 100190, China (e-mail: shuming.tang@ia.ac.cn).

Siji Ma is with the Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China (email:sijima@ieee.org).

Long Chen is affiliated with the State Key Laboratory for Management and Control of Complex Systems at the Institute of Automation, Chinese Academy of Sciences in Beijing 100190, China (e-mail: long.chen@ia.ac.cn). He is also associated with Waytous Co., Ltd. in Beijing, China and the College of Artificial Intelligence at Xi'an Jiaotong University in Xi'an 710049, China.

This work was supported by the National Key Research and Development Program of China (2022YFB4703700); the National Natural Science Foundation of China (Grant No. 62006256, 61773414); Science & Technology Support Project for Young People in Colleges of Shandong Province (Grant No. 2019KJB020).

show bite marks on a biscuit after it's been bitten or accurately represent a chair as a rigid object, leading to unrealistic interactions. Moreover, these models can confuse spatial directions and struggle to depict events over time, such as tracking a camera's movement. These limitations hinder their current application in scenario engineering and end-to-end autonomous mining. However, we expect ongoing technological progress to address these issues soon.

Safety Concerns. Despite advancements in visual fidelity, doubts linger regarding the reliability of these generative models, especially in safety-sensitive domains such as mining. Incorporating potentially flawed data from low-quality video outputs could result in significant real-world consequences. This underscores the urgent need for data review standards before integrating them into autonomous mining operations is emphasized.

II. "6S" MINING WITH IMAGINATIVE INTELLIGENCE AND PARALLEL INTELLIGENCE

The concept of "6S" mining, which prioritizes safety, security, sustainability, sensitivity, service, and smartness, represents the future objectives of the mining industry aligned with sustainable development goals and Environmental, Social, and Governance (ESG) principles [2]. The Sora-like models are powerful tools for simulation and creativity, demonstrating the high-level intelligence of imaginative intelligence [3]. On the other hand, parallel intelligence [4] employs decentralized design principles to empower participants to autonomously simulate various decision paths in a leaderless, trustless environment. These two forms of intelligence either align or have the potential to complement each other. This letter delves deeper into the integration of imaginative intelligence with parallel intelligence to address the aforementioned challenges and advance the "6S" mining initiative.

Three Worlds, Three Types of Miners. Drawing from parallel intelligence and CPSS [5], [6], mining operations can be categorized into three realms: the physical world, involving actual mining and robotic operations; the psychological world, representing the knowledge domain of mining; and the artificial world, consisting of manmade environments such as simulations. These realms correspond to three types of miners: robotic, digital, and human. The advancement of autonomous mining fosters collaboration among these miners, evolving from traditional "human-human collaboration" to more sophisticated "human-machine collaboration" and ultimately to "virtual-reality collaboration" [7]. This new mining paradigm aims to drive the mining industry towards achieving the '6S' goals by integrating AI within the parallel intelligence and CPSS framework.

Alleviating AI Hallucinations. To address AI hallucinations, parallel intelligence presents a promising strategy through computational experiments and the integration of parallel execution within artificial systems. Taking the utilization of a Sora-like model in end-to-end autonomous mining as a case in point, numerous computational experiments can be conducted to assess operational effectiveness across varied contexts. Through iterative experimentation and evaluation, both prompts and generation strategies can undergo continuous adjustment and refinement to optimize the model's performance. This iterative process facilitates reflections on model failures, identification of AI hallucinations, and refinement of strategies based on experiential knowledge acquired from computational experiments. Consequently, computational experiments mitigate AI hallucinations while bolstering mining reliability in real-world scenarios.

Artificiofactual Experiments. Large generative models can bolster the capacity and effectiveness of computational experiments, thereby broadening the scope of artificial social modeling through the integration of imaginative intelligence. This pioneering approach, termed artificiofactual experiments [8], harnesses general artificial intelligence technology to facilitate generative reasoning. By incorporating dynamic visualization reminiscent of Sora-like model's scientific animation format, complex concepts, processes, and the effects of various fine-tuning strategies are vividly portrayed. This format substantially enhances the accessibility, interactivity, and comprehensibility of experiments for both researchers and stakeholders, while also enabling deeper insights into experimental outcomes and facilitating improved understanding and decision-making. This advancement in decision-making capabilities can further contribute to driving the "6S" mining initiative.

III. CONCLUSION

This letter has discussed the potential technological implications of video generation models in smart mining, encompassing scenario engineering, scene understanding, foundation models, end-to-end mining, and human-machine interaction. Despite several challenges that must be addressed before practical implementation, we maintain an optimistic outlook on the prospects of these video generation models in enhancing mining operations, owing to their close alignment with human intelligence.

REFERENCES

[1] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4195–4205.

[2] L. Chen, Y. Li, W. Silamu, Q. Li, S. Ge, and F.-Y. Wang, "Smart mining with autonomous driving in industry 5.0: Architectures, platforms, operating systems, foundation models, and applications," *IEEE Transactions on Intelligent Vehicles*, pp. 1–11, 2024.

[3] F.-Y. Wang, Q. Miao, L. Li, Q. Ni, X. Li, J. Li, L. Fan, Y. Tian, and Q.-L. Han, "When does Sora show: The beginning of tao to imaginative intelligence and scenarios engineering," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 4, pp. 809–815, 2024.

[4] J. Li, R. Qin, S. Guan, X. Xue, P. Zhu, and F.-Y. Wang, "Digital CEOs in digital enterprises: Automating, augmenting, and parallel in Metaverse/CPSS/TAOs," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 4, pp. 820–823, 2024.

[5] L. Li, J. Song, F.-Y. Wang, W. Niehsen, and N.-N. Zheng, "IVS 05: New developments and research trends for intelligent vehicles," *IEEE Intelligent Systems*, vol. 20, no. 4, pp. 10–14, 2005.

[6] J. Yang, X. Wang, Y. Tian, and F.-Y. Wang, "Parallel intelligence in CPSSs: being, becoming, and believing," *IEEE Intelligent Systems*, vol. 38, no. 6, pp. 75–80, 2023.

[7] L. Chen, J. Xie, X. Zhang, J. Deng, S. Ge, and F.-Y. Wang, "Mining 5.0: Concept and framework for intelligent mining systems in CPSS," *IEEE Transactions on Intelligent Vehicles*, 2023.

[8] F.-Y. Wang, Q. Miao, L. Li, Q. Ni, X. Li, J. Li, L. Fan, Y. Tian, and Q.-L. Han, "Sora for computational social systems: From counterfactual experiments to artificiofactual experiments with parallel intelligence," *IEEE/CAA Journal of Automatica Sinica*, 2024.