

## INTRODUCTION

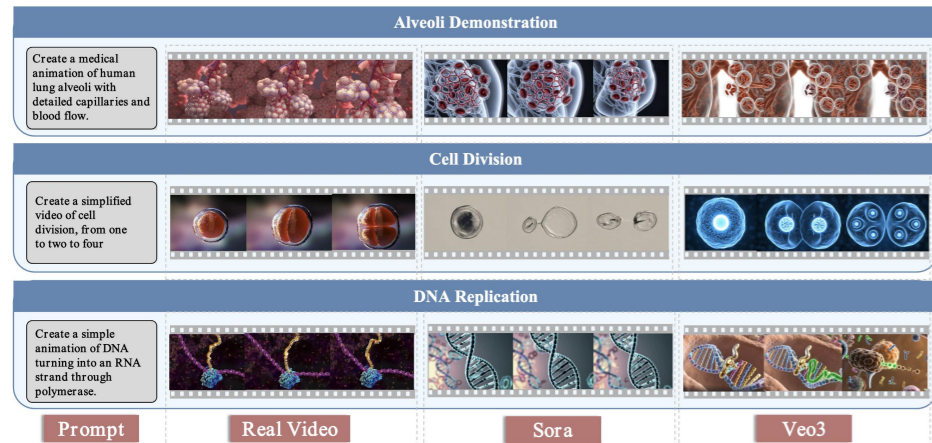


Figure 1. Failure cases of Sora and Veo3 on Microscale Simulation.

### The Rise of World Models.

Recent video generation models like Sora and Veo3 show great promise as world simulators. Trained on vast macroscopic video data, they effectively capture gravity, collisions, and complex object interactions, enabling breakthroughs in robotics, autonomous driving, and embodied AI.

### The Microscopic Gap.

This success has not extended to the microscopic scale. Current models are trained mostly on macroscopic, human-centric data. They lack grounding in biomedical knowledge and microphysical laws. As shown in *Figure 1*, when generating processes like cell division or DNA replication, they produce Scientific Hallucinations: misrepresented fluid dynamics, nonexistent cellular structures, and visually striking but scientifically invalid videos.

### Why Micro Simulation Matters?

Accurate microscale simulation is essential for **biomedical research**, **drug discovery**, **organ on chip systems**, and **STEM education**. Moving beyond costly wet lab experiments demands a new paradigm: models that prioritize scientific correctness over visual plausibility.

## PRELIMINARY EXPLORATION

### ① MicroWorldBench

MicroWorldBench is the first expert-driven benchmark for microscale simulation, enabling systematic evaluation.

► **Hierarchical Scale:** 459 tasks across Organ, Cellular, and Subcellular levels.

► **Multi-Dimensional Scoring:** evaluates Scientific Fidelity, Visual Quality and Instruction Following.

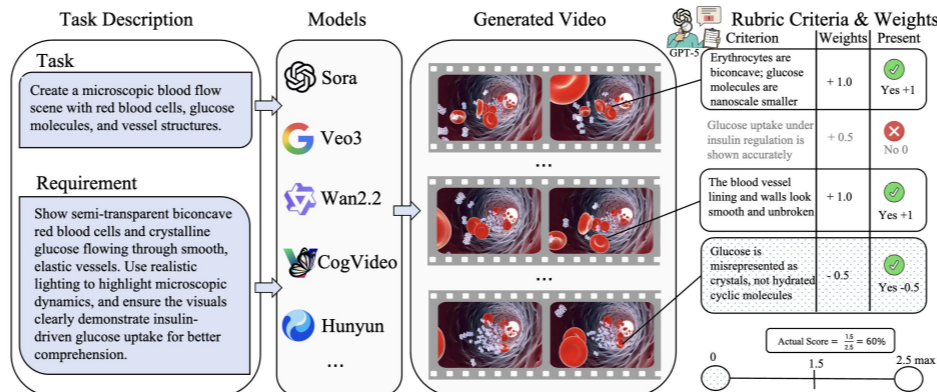


Figure 2. Illustration of MicroWorldBench Evaluation Process.

### ② MicroSim-10K & MicroVerse

We introduce the concept of Micro-World Simulation, driven by world knowledge to ensure the accuracy of microscopic biological mechanisms.

★ **MicroSim-10K (Knowledge-Driven Dataset):** 9,601 expert-verified clips covering organ, cellular, and subcellular dynamics, providing essential physical constraints and biological fidelity.

★ **MicroVerse (Micro-World Model):** Based on MicroSim-10K, it internalizes complex spatiotemporal dynamics and biological rules, accurately reproduces scientific mechanisms.

## RESULTS

### Key Results

- Scaling model size does not improve scientific fidelity.
- Models can mimic organ dynamics via macroscopic intuition, but completely fail at subcellular levels.

Model	Average ↑	Scientific Fidelity ↑	Visual Quality ↑	Instruction Following ↑
<b>Open-Source Video Generation Models</b>				
HunyuanVideo	23.2	15.6	48.2	23.4
CogVideoX-5B	43.5	37.4	64.1	38.6
Wan2.1-T2V-1.3B	49.4	40.3	71.8	50.1
Wan2.2-TI2V-5B	51.6	40.7	82.7	47.0
Wan2.1-T2V-14B	53.8	42.7	86.0	53.8
Wan2.2-T2V-A14B	53.8	37.8	<b>92.8</b>	<b>55.4</b>
<b>MicroVerse-1.3B (Ours)</b>	50.2	<b>43.0</b>	68.5	49.3
<b>Commercial Video Generation Models</b>				
Sora	50.7	35.3	96.4	37.9
Veo3	<b>77.2</b>	<b>65.7</b>	<b>97.0</b>	<b>77.0</b>

Table 1. Performance of different models on MicroWorldBench.

Model	Average ↑	Organ-level ↑	Cellular-level ↑	Subcellular-level ↑
<b>Open-Source Video Generation Models</b>				
HunyuanVideo	23.2	23.1	23.8	19.4
CogVideoX-5B	43.5	39.9	47.0	38.6
Wan2.1-T2V-1.3B	49.4	45.9	51.7	52.4
Wan2.2-TI2V-5B	51.6	46.6	53.9	49.5
Wan2.1-T2V-14B	53.8	55.7	<b>54.4</b>	52.8
Wan2.2-T2V-A14B	53.8	<b>56.3</b>	52.0	53.3
<b>MicroVerse-1.3B (Ours)</b>	50.2	47.6	51.7	<b>53.3</b>
<b>Commercial Video Generation Models</b>				
Sora	50.7	55.9	46.1	55.0
Veo3	<b>77.2</b>	<b>77.5</b>	<b>76.9</b>	<b>78.2</b>

Table 2. Performance (Dimension-Wise) of different models.

## FUTURE WORK

Explicitly incorporate mathematical physical laws and biochemical equations as hard constraints within the model to enhance scientific rigor and transition from visual approximation to precise microscopic world simulation.