

AMCC 5160 • Lecture 09 readings

All three connect directly to tonight's stories: one on generated 3D worlds, one on robots learning from people, and one on how the field talks about world models. For Quiz 09, reflect on any one of the three. Two optional technical readings are listed last.

Reading 1 / Technique: Atlas: a world model for spatial intelligence

World Labs • Company blog | 1 September 2026 — <https://www.worldlabs.ai/blog/atlas>

What to do: Read the sections on the spatial context, camera-controlled generation, reconstruction and robotics, and watch two of the demos. You do not need the benchmark tables.

Summary: Atlas is one model for images, video and 3D. Every input (text, photos, depth, video) is placed into a shared spatial context with explicit camera poses, and the model can output new views, minute-long 1440p camera-controlled video, depth maps, point clouds or Gaussian splats. With two or three ordinary photos it reconstructs a scene; with phone video it can simulate what a robot would see moving through a room. It is in early access, with no paper yet.

Questions to keep in mind:

- What does it change to pin every image to a camera position in 3D?
- Which demo is reconstruction (faithful) and which is generation (invented)? How can you tell?
- World Labs gives only its own benchmark numbers. What would you test first?

Connection: Story 2: “Atlas: one model for pictures, video and 3D” and “Atlas, by World Labs’ own numbers” .

Context: AMD to acquire World Labs (28 Sept 2026) —

<https://ir.amd.com/news-events/press-releases/detail/1299/amd-to-acquire-world-labs-to-advance-the-future-of-ai-compute>

Reading 2 / Robotics: $\lambda 0$: a humanoid foundation model that learns whole-body tasks from human data

Alibaba Amap and Token Hub • Project page | October 2026 — <https://lambdahumanoid.github.io/>

What to do: Watch the report video, then read the parts on HumanVerse (the capture setup), the three training stages, and the ablation results.

Summary: $\lambda 0$ is a vision-language-action model for a Unitree G1 humanoid with five-fingered hands. It is trained in three stages: first-person human hand data, then 500 hours of whole-body human capture (HumanVerse) translated into the robot's view and joints, then 600 real robot demonstrations. It succeeded in 25 of 40 real trials (62.5%) versus 40% for $\pi 0.5$; without the human data, success fell to 15%.

Questions to keep in mind:

- Why does the capture camera sit on the neck, at the robot's head height?
- What does the ablation tell you about where the robot's skill comes from?
- Would you wear the HumanVerse rig for a week? What would you refuse to record?

Connection: Story 3: $\lambda 0$ slides and the discussion on whose daily lives become training data.

News coverage: TechNode, 26 Oct 2026 — <https://technode.com/2026/10/26/alibabas-amap-unveils-humanoid-model-that-learns-whole-body-tasks-from-human-movements>

Reading 3 / Culture: DeepMind thinks its new Genie 3 world model presents a stepping stone toward AGI

Rebecca Bellan, TechCrunch • 5 August 2025 — <https://techcrunch.com/2025/08/05/deepmind-reveals-genie-3-a-world-model-that-could-be-the-key-to-reaching-agi>

What to do: Read the article (about 6 minutes). Note every claim DeepMind makes about what world models are for, and every limitation the article mentions.

Summary: DeepMind presents Genie 3 as the first real-time, general-purpose interactive world model: minutes of 720p, 24 fps worlds from a text prompt, with “promptable world events” and a memory of what it generated. The team argues its main use is training general-purpose agents, and so a step towards AGI. The article also notes limits, such as unrealistic snow physics in one demo and a narrow range of actions.

Questions to keep in mind:

- Why does DeepMind say memory matters more than image quality?

- Is training agents a different goal from making art or games? Where do they overlap?
- Use this article in tonight's debate: does it support "a video generator is a world model" ?

Connection: Story 1 and the debate in Story 4.

Official post — <https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/>

Optional / Technique: Gemini 4 Argon for robotics

Google for Developers • 13 October 2026 — <https://developers.googleblog.com/gemini-4-argon-for-robotics/>

How a general model writes MuJoCo simulations from video and sketches, and uses them to train and improve robots.

Optional / Technique: Gaussian Splatting for Novel View Synthesis (lecture slides)

Stanford CS231A, Lecture 17 • Slides — https://web.stanford.edu/class/cs231a/lectures/lecture17_gaussian_splatting.pdf

The full version of tonight's NeRF → 3DGS slides, including the projection maths.