Human-Computer Interaction

XR & Spatial Computing

Professor Bilge Mutlu

Today's Agenda

- Topic overview: *XR & Spatial Computing*
- The frame: *enduring fundamental → how AI changes it*
- Provocation: *Does AI re-found XR, or only its content supply?*
- Discussion

This term's frame

Enduring fundamental → How AI is changing it → Depth (a claim to argue with)

- **Today's fundamental:** *presence*, place illusion + plausibility illusion (Slater): a fact about human perception, not hardware
- **How AI changes it:** it re-founds the **content & authoring** pipeline, *not* the perceptual/interaction core
- **Depth: Moderate.** AI re-founds the *supply of content*, not what XR fundamentally *is*

This week's readings

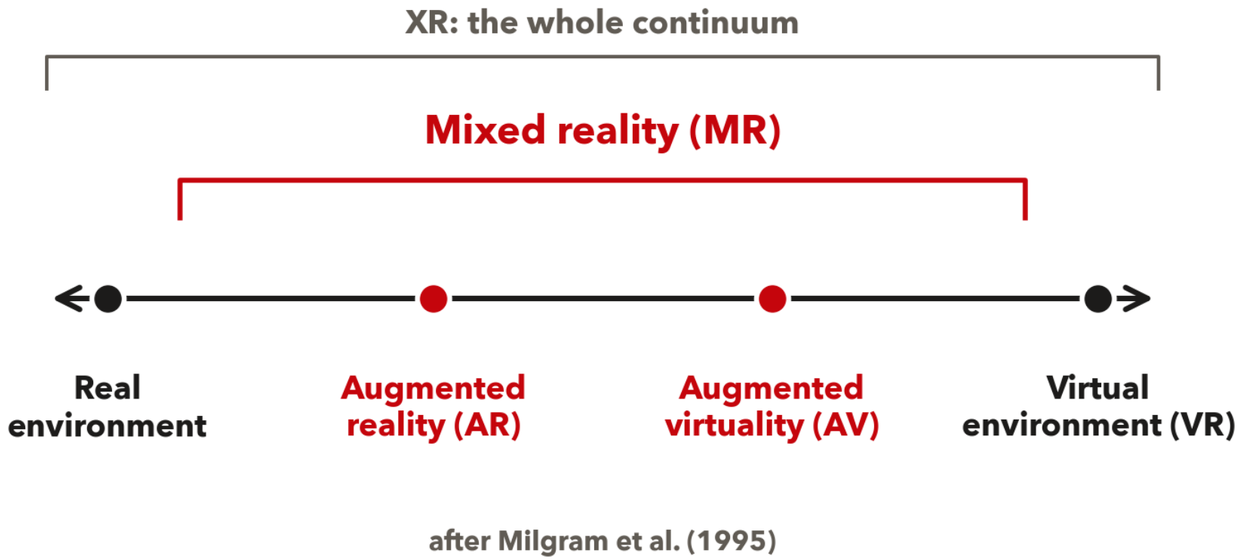
- **Required**
  - Slater, *Place Illusion and Plausibility* (2009)
  - Zhang et al., *VRCopilot* (UIST 2024)
- **Recommended**: Zhu, Chen & Li, *When Generative AI Meets Extended Reality* (2026)
- **Optional**: Slater et al., *A Separate Reality* (2022) · Fitts (1954)

Topic overview: *XR & Spatial Computing*

What is XR? [1]

XR spans the reality–virtuality continuum: VR, AR, MR.

- **VR:** a computer-generated world: 3D, interactive, tracked to the user
- **AR:** (1) combines real + virtual, (2) interactive in real time, (3) *registers* virtual imagery with the real world; supplements rather than replaces



[1] Azuma, R.T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, 6(4), 355–385. Continuum: Milgram, P., Takemura, H., Utsumi, A. & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Proc. SPIE* 2351, 282–292.

The enduring fundamental: presence [2]

Mel Slater: presence is not a property of the *display*; it is elicited from the *user's* perceptual-cognitive system.

Place illusion (PI)

"I am *here*": the sensorimotor feeling of being in the place

Plausibility illusion (Psi)

"This is *happening*": events appear to really be occurring

[2] Slater, M. (2009). Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B*, 364(1535), 3549–3557.

Why presence is the fundamental

Presence is elicited, not rendered: a fact about human perception, not about how the content was produced.

- When PI and Psi hold, people respond **realistically**
- This is what XR *is*; it predates and survives any model
- *A Separate Reality* (2022) [3]: the account holds however the content is made

[3] Slater, M., et al. (2022). A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality*, 3.

The perceptual/interaction core

The hard problems AI does **not** re-found, because they are about the human perceptual system:

- **Registration:** virtual content locked to the real world
- **Latency:** the motion-to-photon budget
- **Vergence–accommodation conflict:** the eyes' unresolved depth cue
- **Cybersickness:** the body's response to mismatch

But what about *AI*?

AI re-founds the *content* pipeline [4]

World-authoring collapses from **artist-months → prompt-seconds**:

- **World models:** explorable environments from a prompt
- **Text/image → 3D:** assets and scenes from a prompt
- **Gaussian splatting:** capture reality, replay it



[4] **Genie 4**: Google DeepMind, Oct 27 2026; video from its announcement, `deepmind.google`. **Marble**: World Labs, launched Nov 12 2025, `worldlabs.ai`. Company announcements, not peer-reviewed research.

The industry claim [5]

"widespread XR adoption remains limited due to two key challenges: 1) the high cost and complexity of authoring 3D content... and 2) the steep learning curve associated with non-intuitive interaction methods"

- Generative AI is offered as the fix for *both*: "automating content generation" and "language-driven interaction"

[5] Zhu, M., Chen, J. & Li, B. (2026). When Generative AI Meets Extended Reality: Enabling Scalable and Natural Interactions. *IEEE Internet Computing*. arXiv:2601.15308. *(A magazine article; read it critically.)*

Fix for whose problem?

- Challenge 1 is the **builder's** problem; challenge 2 is the **user's**, often a different person
- Generated content helps the builder; the user still has to point, select, and move
- Language input does not change that: the **interaction core** is untouched

Two artifacts. They **disagree**.

Genie 4 • DeepMind

Real-time worlds with native audio and dialogue, steered by prompt as they run. Lasts **a few minutes**; details and object permanence drift.

There is no scene. Only the next frame.

Marble • World Labs

Text, images, video, or a coarse 3D layout → a **persistent** world: Gaussian splats *plus collider and high-quality meshes* for standard 3D tools.

Video is just one export.

What each one buys

Map them onto Slater and the disagreement resolves:

- **Marble buys place illusion:** a stable *somewhere*; turn around and it is still there
- **Genie buys plausibility illusion:** *everything* responds, instantly, including what nobody authored
- Each purchases one half of presence by giving ground on the other

Neither is "better." They are optimizing different halves of Slater.

The evidence: automation costs agency [6]

"Scaffolded creation using wireframes enhanced the user agency compared to automatic creation… manual creation via multimodal specification offers the highest sense of creativity and agency."

- The **most automated** condition lost on agency
- Reported by authors whose own contribution *is* generative AI in VR
- Wireframes, the paper's **intermediate representation**, hand layout control back to the author

---

[6] Zhang, L., Pan, J., Gettig, J., Oney, S. & Guo, A. (2024). VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. *UIST '24*. doi:10.1145/3654777.3676451 · arXiv:2408.09382.

The instructive line

AI re-founds the supply of content, without re-founding what XR fundamentally is.

- The **content pipeline** is transformed (author worlds in seconds)
- The **perceptual/interaction core** is untouched (presence, registration, latency, VAC, sickness)
- Immersion/fidelity is *engineered*; presence is a property of the **user's** system: *elicited, not rendered*

The boundary to teach

	Week 8: Ubicomp *(Light)*	**Week 10: XR *(Moderate)***
What AI touches	sharpens *sensing/inference*	re-founds *content/authoring*
Fundamentals	Weiser's questions survive	presence survives
Verdict	AI does *not* re-found it	AI re-founds the *pipeline*

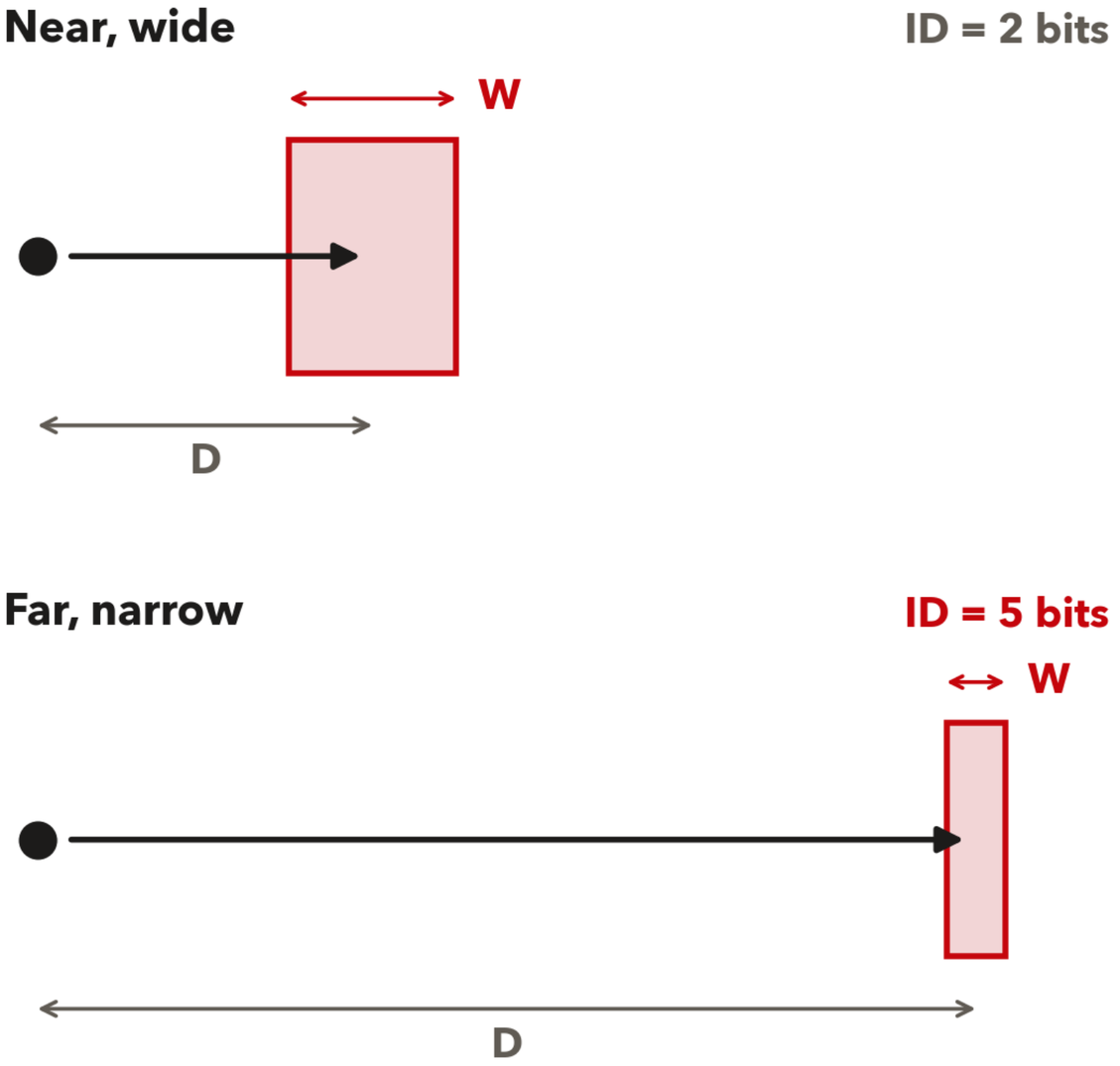
The single most instructive boundary of the term: AI re-founds the **production pipeline** (XR) vs. does **not** re-found the **interaction fundamentals** (Ubicomp).

Fold-in: Fitts's Law in 3D [7]

The durable motor model for pointing and selection:

$$MT = a + b \log_2 \frac{2D}{W}$$

- *MT* movement time, *D* distance, *W* target width
- Log term: the **index of difficulty** (bits)
- *a*, *b* fit per device; holds in 3D and mid-air



[7] Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.

Guardrails coda

- **Immersive persuasion:** presence makes people respond realistically, so a persuasive scene acts as an experience, not a message
- **Whose reality gets authored:** a generated world reflects the builder's prompt and the model's training data; the user lives inside it
- **Spatial-sensing data:** headsets track eyes, hands, body, and room just to work; 100 seconds of head and hand motion picked out one user among 50,000+ (94%) [8]

[8] Nair, V., Guo, W., Mattern, J., Wang, R., O'Brien, J. F., Rosenberg, L. & Song, D. (2023). Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data. *USENIX Security '23*, 895–910. arXiv:2302.08927.

Your provocation *(post one slide)*

Pick one:

1. **Provocation:** a defensible claim about whether AI-generated worlds change what XR *fundamentally is* (presence, perception) or only its *content supply*, grounded in Slater.
2. **Critical artifact:** take **Genie 4** or **Marble** (or another generative-3D / world-model demo); annotate where it touches the *perceptual/interaction core* vs. only the *content pipeline*.
3. **Design / policy:** one principle for XR that treats presence as *elicited from the user's perceptual system*, not *rendered* by better content.

Discussion

Today's roles

Reporter	Skeptic	Connector
Seat 2	**Seat 3**	**Seat 4**

Same in every pod. **Seats 1, 5 and 6** have no role today; you are still in the argument.

Discussion Format

In your pod, on your pod monitor: ~30 minutes.

- **Everyone shows their slide**, about 2 minutes each, cast from your own laptop
- **Skeptic** puts a counterargument on the table · **Connector** ties it to an earlier week
- **Reporter** writes the pod's three lines (agreed, split, unresolved) and delivers them to the room

What's Next?

- **Friday, Methods 9: Statistics II** (Nov 6)
  - **Read:** the textbook **Chapter 4**, inferential tests + regression
  - **Due 9:30:** your *Descriptive Summary* + *Effect Sizes* write-up
  - Team **project status** on Canvas by 9:30 (done / blocker / next)