

When Direct Manipulation Becomes a Guess: Productive Friction in AI-Mediated Multisensory Visualization

Anchit Mishra

University of Waterloo

ABSTRACT

Generative visualization increasingly embeds large language models within direct manipulation and multisensory interaction. Speech, gaze, touch, gesture, sound, and haptics can make probabilistic inference feel like familiar, deterministic tool use. I call this gap a *deterministic-affordance mismatch*: deterministic interaction cues persist while AI weakens predictability, locality, reversibility, or provenance. Reading the malleable interfaces of *Iron Man 2* against systems from *Blade Runner 2049*, I propose *productive friction*, including cross-sensory renderings whose detail reflects model uncertainty. Four frictions expose the inference boundary, match action to effect, make stochastic branches tangible, and attribute sensory agreement.

Index Terms: Generative Visualization, Multisensory Interaction, Direct Manipulation, Human–AI interaction.

1 INTRODUCTION

In an *Iron Man 2* scene, Tony Stark speaks, points, and expands a projected city model while JARVIS supplies the missing computation [7]. The artefacts persist, yet an intelligent agent interprets each incomplete action. Current generative visualization resembles this fiction. DynaVis turns language into persistent widgets [16]; Data Formulator delegates transformation to an AI agent [18]; and InterChat combines direct manipulation with LLM-based intent inference [3]. They motivate a question: when does a probabilistic model adopt a deterministic tool’s perceptual grammar?

Science-fiction interfaces act as diegetic prototypes shaping technological expectations [9]. They expose assumptions technical accounts leave implicit [4]. I identify this mismatch and four productive frictions that make consequential mediation perceptible.

2 A DIRECT ACTION IS A PROMISE

Direct manipulation rests on continuous representation, physical action, rapid feedback, and incremental, reversible operations [15]. Felt directness also depends on short semantic and articulatory distances between intention, action, and result [8]. Multimodal systems complicate this contract. They combine complementary signals, using one mode to resolve another’s ambiguity [12]. A gaze can identify an object while speech names an operation.

LLM mediation can weaken this contract at every stage in Fig. 2. A small gesture may trigger global changes. The same action may yield another result after context, sampling, or model updates. An apparent undo may regenerate instead of restoring an exact state. Fast feedback can hide these breaks because the action still *feels* local, causal, and repeatable.

Multisensory output can strengthen that feeling. People integrate visual and haptic estimates by apparent reliability [5]; haptics can encode data through force, texture, or vibration [13]. A single model can coordinate graphics, narration, and haptic emphasis,

making one error feel like several reinforcing sensations. Exact-looking marks, tones, or haptic events can also overstate a coarse inference. Designers can use purposefully low-fidelity sensory representations to communicate uncertainty in model output. A coarse output could appear as a visual interval, a sonified band, or a diffuse haptic region. Each replaces an exact mark, pitch, or detent. The representation’s perceptual resolution should match the inference’s warranted resolution. Exact source measurements should remain available for inspection. No sense is inherently low fidelity; designers choose the detail encoded within it. The mismatch therefore has four failure points: hidden inference weakens predictability; disproportionate effects weaken locality; regeneration weakens reversibility; and shared multisensory generation obscures provenance.

3 TWO LESSONS FROM *Blade Runner 2049*

Blade Runner 2049 offers an operational counterpoint to JARVIS [17]. Its Denabase presents genomic and birth records through a mechanical, microfilm-like browser. K advances and compares records until a duplicated pair appears. In the morgue, dials and optical transitions expose increasing magnification from bone to microscopic serial number. The design team describes both systems through physical, optical, microfiche, and mechanical references [6]. Their friction makes evidence, comparison, and analytical labour visible. Material controls provide no inherent truth, but these scenes keep intermediate states and the route from specimen to finding inspectable.

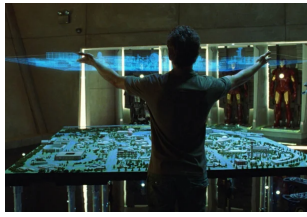
Joi’s holographic artefacts disclose dependence on projectors and infrastructure. Replicants possess material bodies, yet embodiment cannot settle their personhood. The Joi–Mariette encounter distributes visual presence, touch, speech, and agency across several sources. Productive friction should disclose provenance and AI mediation when aligned sources might otherwise merge. It should never become a sensory test for deciding who, or what, deserves recognition. The goal resembles Wheeler’s “reappearing tool”: mediation should become visible when transparency creates epistemic risk [19].

4 FOUR FRICTIONS FOR FOUR WEAKENED PROMISES

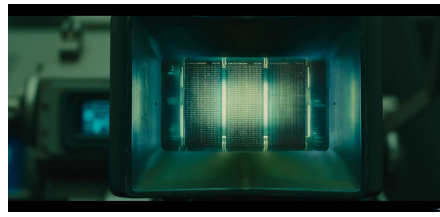
Each friction answers one failure point in the deterministic-affordance mismatch. MacLean describes haptics as making digital transactions physical and confirmed [10]; I extend this principle to mark the boundary between direct input and model inference.

1. **Predictability**—expose inference. JARVIS completes Stark’s partial speech and gesture. A real interface should expose that handoff before execution, consistent with human–AI guidance [1]. It can ghost the inferred target, operation, and cross-modal binding; a haptic detent, auditory transition, or visual threshold can separate exploration from commitment.

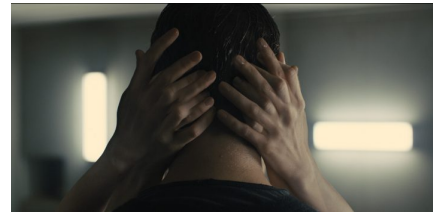
2. **Locality**—match action to effect. A small gesture should not silently trigger a global transformation. Before execution, a scope halo should reveal every affected dataset, view, and encoding. Effects exceeding the action’s apparent scale require an ex-



(a) Malleable model



(b) Inspectable records



(c) Mixed embodiment

Figure 1: Three science-fiction scenes expose different interface promises. (a) JARVIS interprets Stark’s expansive gesture over a malleable model. (b) A backlit Denabase card keeps record comparison physically visible. (c) Duplicated hands in the Joi–Mariette merge disclose multiple embodied sources. Panels (a–b) are visualization cases; panel (c) is an ethical analogy.

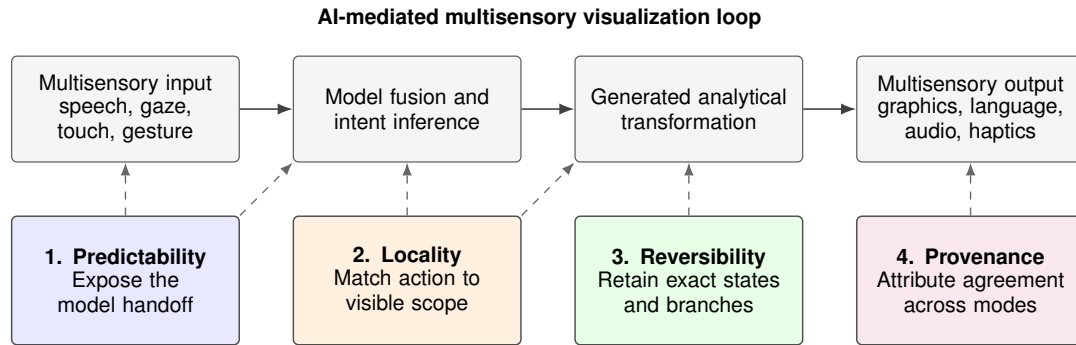


Figure 2: An AI model can mediate the full multisensory visualization loop. Each friction answers a direct-manipulation promise weakened by model mediation: predictability, locality, reversibility, or provenance. At the output stage, sensory detail can also reflect model uncertainty.

licit second commitment. An inspectable trace exposes generated code, assumptions, and downstream changes.

3. **Reversibility**—make branches tangible. The Denabase and morgue sequences keep intermediate evidence visible as K advances. Generative interfaces should likewise preserve exact checkpoints and expose alternative histories. Undo restores the prior state; replay reuses captured inputs and settings; regenerate creates a labelled stochastic branch. Persistent visual, auditory, or tangible tokens can support comparison between histories.

4. **Provenance**—attribute sensory agreement. The Joi–Mariette analogy keeps several embodied sources visible despite aligned outputs. AI interfaces should likewise attribute agreement across graphics, narration, sonification, and haptic emphasis. Co-ordinated cues need a common-source marker to avoid implying independent corroboration. At the same time, measured data, external evidence, and model inference should remain separable.

5 CONCLUSION

Productive friction must remain selective. Cognitive forcing can reduce overreliance but may be disliked [2]. Designers should scale both friction and cross-sensory detail with uncertainty, consequence, and reversibility. Every cue needs an accessible alternative; haptics should support access and embodiment without becoming mandatory evidence.

This paper is intended to provoke further discussion and debate rather than present an evaluated design framework; it raises two empirical questions:

1. How should sensory detail scale with model uncertainty?
2. How should friction scale with an operation’s scope and reversibility?

As visualization becomes more malleable and perceptually fluent, designers should dedicate effort to preserving directness (and thus, perceived user agency), expose AI limits, and prevent sensory detail from overstating model certainty. This way, interfaces of the future can make use of human sensory perception in a way that conveys AI mediation more realistically instead of exploiting multisensory perception to reinforce weak, uncertain, or inaccurate claims.

FIGURE CREDITS

Figure 2 was created by the author for this paper. Figure 1 uses film frames for scholarly criticism and analysis. Panel (a): *Iron Man 2* (2010), © 2010 MVL Film Finance LLC; source: Ortiz [11]. Panels (b–c): *Blade Runner 2049* (2017), © 2017 Alcon Entertainment, LLC; sources: Territory Studio via Failes [6] and fxguide [14]. Composition and labels are by the author.

REFERENCES

- [1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human–ai interaction. In *Proc. CHI*, pp. 1–13. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300233 1
- [2] Z. Bućinca, M. B. Malaya, and K. Z. Gajos. To trust or to think: Cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–21, 2021. doi: 10.1145/3449287 2
- [3] J. Chen, J. Wu, J. Guo, V. Mohanty, X. Li, J. P. Ono, W. He, L. Ren, and D. Liu. InterChat: Enhancing generative visual analytics using multimodal interactions. *Computer Graphics Forum*, 44(3):e70112, 2025. doi: 10.1111/cgf.70112 1
- [4] P. Dourish and G. Bell. “resistance is futile”: Reading science fiction alongside ubiquitous computing. *Personal and Ubiquitous Computing*, 18(4):769–778, 2014. doi: 10.1007/s00779-013-0678-7 1

- [5] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a [1](#)
- [6] I. Failes. A visual journey through the screen graphics of Blade Runner 2049. vfxblog, with production notes from Territory Studio, 2017. Accessed August 20, 2026. [1](#), [2](#)
- [7] J. Favreau. Iron Man 2. Motion picture, Marvel Studios and Paramount Pictures, 2010. Director. [1](#)
- [8] E. L. Hutchins, J. D. Hollan, and D. A. Norman. Direct manipulation interfaces. *Human–Computer Interaction*, 1(4):311–338, 1985. doi: 10.1207/S15327051HCI0104_2 [1](#)
- [9] D. A. Kirby. The future is now: Diegetic prototypes and the role of popular films in generating real-world technological development. *Social Studies of Science*, 40(1):41–70, 2010. doi: 10.1177/0306312709338325 [1](#)
- [10] K. E. MacLean. Using haptics for mobile information display. In *Proc. Pervasive Mobile Interaction Devices (PERMID)*, pp. 175–179. Sydney, Australia, 2008. [1](#)
- [11] Z. Ortiz. Iron Man II: JARVIS hologram sequence. Design portfolio, 2020. Accessed August 20, 2026. [2](#)
- [12] S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, 1999. doi: 10.1145/319382.319398 [1](#)
- [13] S. Panëels and J. C. Roberts. Review of designs for haptic data visualization. *IEEE Transactions on Haptics*, 3(2):119–137, 2010. doi: 10.1109/TOH.2009.44 [1](#)
- [14] M. Seymour. The techniques used in the Blade Runner 2049 hologram sex scene. fxguide, 2017. Accessed August 20, 2026. [2](#)
- [15] B. Shneiderman. Direct manipulation: A step beyond programming languages. *Computer*, 16(8):57–69, 1983. doi: 10.1109/MC.1983.1654471 [1](#)
- [16] P. Vaithilingam, E. L. Glassman, J. P. Inala, and C. Wang. DynaVis: Dynamically synthesized ui widgets for visualization editing. In *Proc. CHI*. ACM, New York, NY, USA, 2024. doi: 10.1145/3613904.3642639 [1](#)
- [17] D. Villeneuve. Blade Runner 2049. Motion picture, Alcon Entertainment and Warner Bros. Pictures, 2017. Director. [1](#)
- [18] C. Wang, J. Thompson, and B. Lee. Data formulator: AI-powered concept-driven visualization authoring. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1128–1138, 2024. doi: 10.1109/TVCG.2023.3326585 [1](#)
- [19] M. Wheeler. The reappearing tool: Transparency, smart technology, and the extended mind. *AI & Society*, 34(4):857–866, 2019. doi: 10.1007/s00146-018-0824-x [1](#)