Directing the Robot: Scaffolding Creative Human–AI–Robot Interaction

Jordan Aiko Deja
De La Salle University
Manila, Philippines
jordan.deja@dlsu.edu.ph

Isidro Butaslac III
Nara Institute of Science and Technology
Ikoma, Japan
isidro.b@naist.ac.jp

Nicko Reginio Caluya
Ritsumeikan University
Osaka, Japan
nicko@fc.ritsumei.ac.jp

Maheshya Weerasinghe
University of Primorska
Koper, Slovenia
maheshya.weerasinghe@famnit.upr.si

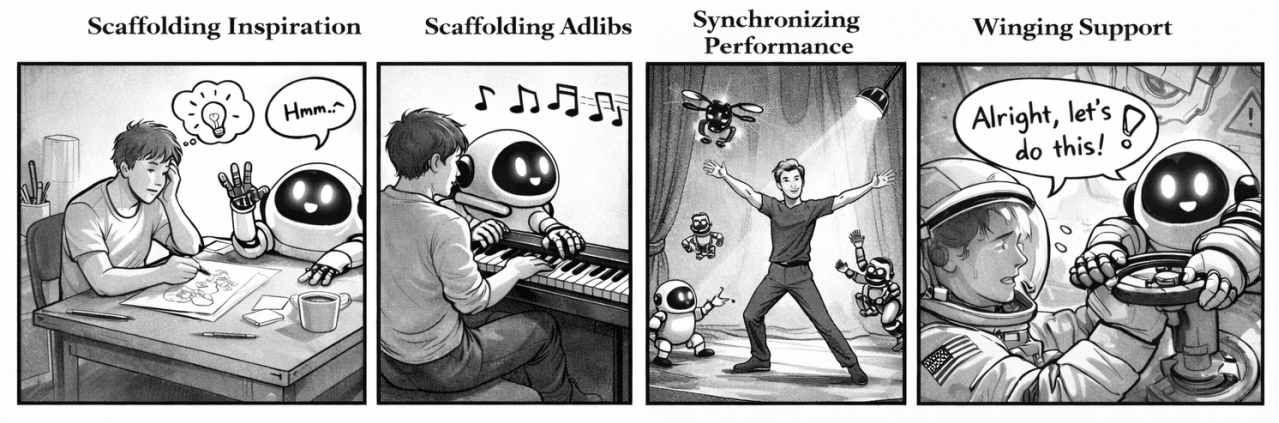

Figure 1: Illustrative scenarios of scaffolding in creative human–AI–robot interaction. From left to right: (1) Scaffolding Inspiration (robot gestures support ideation); (2) Scaffolding Adlibs (robot improvises in musical performance); (3) Synchronizing Performance (ground and aerial robots in stage collaboration); and (4) Winging Support (robot assistance under time pressure). Across scenarios, AI scaffolds creativity and action while preserving human agency. ***Illustration generated using generative AI and refined by the authors.***

Abstract

Robots are moving beyond industrial settings into creative, educational, and public environments where interaction is open-ended and improvisational. Yet much of human–AI–robot interaction remains framed around performance and efficiency, positioning humans as supervisors rather than collaborators. We propose a re-framing of AI interaction with robots as scaffolding: infrastructure that enables humans to shape robotic behaviour over time while remaining meaningfully in control. Through scenarios from creative practice, learning-by-teaching, and embodied interaction, we illustrate how humans can act as executive directors, defining intent and steering revisions, while AI mediates between human expression and robotic execution. We outline design and evaluation implications that foreground creativity, agency, and flow. Finally, we discuss open challenges in social, scalable, and mission-critical contexts. We invite the community to rethink interacting with Robots and AI not as autonomy, but as sustained support for human creativity.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction design*; • **Computing methodologies** → *Robotics*; *Artificial intelligence*.

Keywords

robotics, drones, human-robot interaction, scaffolding, creativity, mission critical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
The 3rd InterAI Workshop at CHI 2026, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Jordan Aiko Deja, Isidro Butaslac III, Nicko Reginio Caluya, and Maheshya Weerasinghe. 2026. Directing the Robot: Scaffolding Creative Human–AI–Robot Interaction. In *The 3rd InterAI Workshop: "Interactive AI for Human-Centered Robotics at ACM CHI 2026, April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction and Background

Robots are increasingly becoming part of daily life, expanding beyond industrial settings [10] into public spaces, educational environments [18], and creative contexts [19]. As robots move closer to people in both physical and social settings, interaction with them is no longer exceptional but inevitable. Advances in Artificial Intelligence (AI) have been central to this shift, often by making human–robot interaction more accessible through the automation of behaviours, optimisation of performance, and reduction of human effort in repetitive or routine tasks. While effective, these approaches have also shaped dominant design paradigms that prioritise efficiency and autonomy over richness of intent.

However, framing AI primarily as a tool for automation and optimisation limits how humans and robots can collaborate in open-ended settings. In such framings, humans are often positioned as supervisors or end-users, while decision-making authority and interpretive control are increasingly delegated to the system. In this context, interpretative control is defined as how the human decides the execution similar to how the director of photography decides on the color toning and angles when shooting a film. This diminishes the value of human autonomy [7], and limits active engagement, exploration, and creative collaboration, particularly in expressive, educational, or improvisational contexts.

Prior work has explored improvisation, mixed-initiative interaction, and co-creative human–AI systems [1, 5, 11, 13]. However, scaffolding in these systems is often treated as either a technical mechanism for learning or a temporary design strategy used to compensate for system limitations. In contrast, we argue for treating scaffolding as a "core interaction principle" in human–AI–robot collaboration; one that explicitly structures how agency, interpretation, and control are distributed over time. In this case, we look at not leaving it all to the AI or robot, but rather having the original idea come from the human and both human and AI collectively build on their ideas.

Building on this perspective, we propose a re-framing in which interactive AI scaffolds collaboration by positioning humans as "executive directors" of robotic behaviour. In this role, humans define intent, make interpretive judgments, and steer iterative revisions, while interactive AI mediates between human expression (e.g., language, gesture, demonstration) and robotic execution (e.g., motion, coordination, embodied action). Rather than handing control away from the human, AI supports thinking, decision-making, and creative flow, while robots act as expressive and physical performers of human-defined intent. This perspective foregrounds human agency and re-frames interaction as an ongoing creative process rather than a supervisory task. Specifically, this paper contributes (1) a re-framing of interactive AI as scaffolding for creative human-robot interaction, (2) illustrative scenarios spanning creative and mission-critical contexts (as shown in Figure 1), and (3) design and evaluation implications centered on agency and flow.

2 Rethinking Design and Evaluation Through Use Cases

To ground the idea of scaffolding as an interaction principle, we consider use cases in which creativity, exploration, and improvisation are central. In many such contexts, robots are not continuously autonomous actors, but remain available as expressive resources that can be activated, redirected, or reinterpreted by humans as needed. Interactive AI mediates this availability, allowing humans to focus on intent, timing, and meaning rather than low-level control. In this sense, interaction unfolds as a process of direction rather than execution. Across these scenarios, humans act as executive directors: they articulate high-level goals, shape interpretation, and intervene iteratively as outcomes emerge. Interactive AI supports this role by translating underspecified human input into executable robotic behaviour, managing ambiguity and coordination while leaving authorship and decision-making with the human.

> **Thought 1: Preserve interpretive control.** Interactive AI must support human intent without displacing authorship or the ability to redirect action.

One illustrative case involves teaching or shaping robot motion through few-shot demonstrations [1]. A human may imagine a sequence of movements or expressive qualities they wish a robot to perform, but may be constrained by physical ability, scale, or precision. Rather than requiring exact replication, interactive AI can scaffold this process by interpreting partial demonstrations, smoothing transitions, and generating feasible motion variants that preserve the intended structure and emphasis of the human (Figure 1 Frame 1). Here, the human remains responsible for defining what the movement should convey, while AI manages how that intent is realized in robotic execution.

Also as in emerging learning-oriented contexts, where humans do not merely demonstrate tasks, but iteratively shape how robots collaborate with them over time. For instance, in collaborative assembly settings, researchers have developed a "Teaching-Learning-Collaboration (TLC)" model in which a human demonstrates a task and communicates high-level intent and preferences to a robot collaborator [25]. The robot uses this information to build task knowledge via learning from demonstrations and then actively assists the human in completing shared work. Such models illustrate how humans can act as executive directors: guiding, teaching, and co-revising robotic behaviour while interactive learning enables robots to translate human intent into embodied collaborative action without displacing control.

These examples suggest several design implications for interactive AI as scaffolding. First, systems should prioritise "responsiveness over correctness", enabling interaction to continue even when human input is partial, ambiguous, or evolving. Second, interfaces that leverage familiar forms of expression such as language, gesture, or embodied demonstration, can reduce cognitive overhead by allowing humans to direct robots in ways that align with how they already communicate intent. In such cases, AI acts less as an autonomous decision-maker and more as a translator between human expression and robotic performance.

> **Thought 2: Prioritise responsiveness over optimisation.** Interactive AI should sustain creative flow under evolving input rather than converging on a fixed optimal solution.

This framing is particularly relevant in domains such as human–drone and multi-robot interaction, where direct manual control

is often impractical. In performative or large-scale settings, such as stage productions involving drones or zoomorphic robots [16, 17], humans may cue, redirect, or improvise in real time using gestures or brief verbal instructions. Interactive AI can scaffold coordination and timing across multiple robots (Figure 1 Frame 3), enabling collective behaviour that remains responsive to human direction rather than pre-scripted autonomy. Shape, motion, and embodiment further support imagination and storytelling, allowing robots to function as expressive performers rather than independent agents [15, 24].

Rethinking design this way also requires rethinking evaluation. Traditional HRI metrics such as efficiency, error rate, or task completion time are poorly suited to interactions centered on exploration and improvisation. Instead, evaluation should ask whether systems support humans in experiencing themselves as directing the interaction–able to reinterpret, intervene, and steer outcomes as they unfold. Metrics that capture flow [4], perceived agency, and sustained engagement, such as the Creativity Support Index [3], provide more appropriate lenses for assessing whether interactive AI functions effectively as scaffolding rather than as automation.

> **Thought 3: Measure agency as a first-class outcome.** Evaluate scaffolding by users' sense of authorship and flow, not just efficiency.

Framing evaluation around creativity and agency also serves a broader purpose. By explicitly valuing human direction and interpretive control, interactive AI systems resist narratives of robots replacing or overpowering human collaborators. We can do this by designing systems that support humans in articulating intent, maintaining creative flow, and remaining accountable for outcomes sometimes through something as simple as a gesture, a cue, or a brief corrective intervention [21]. In this way, evaluation becomes a design commitment rather than a post-hoc measurement.

3 Open Challenges and Broader Implications

We also raise a set of open challenges that extend beyond individual use cases. One challenge concerns social interaction and physical proximity in close human–robot collaboration [20]. When humans direct robots through naturalistic cues such as gesture, touch, or expressive movement, the boundaries of control and consent may become ambiguous. In such interactions, responsibility does not lie solely in issuing commands, but in shaping interpretation and responding to system behaviour as it unfolds. Designing interactive AI that can scaffold expressive direction while making the boundaries of agency legible remains an open problem.

Another challenge arises when scaling the direction from a single robot to multiple robots or swarms. As humans gesture, speak, or demonstrate intent to multiple robots simultaneously, how should that intent be preserved, transformed, or distributed throughout the system [8, 23]? Although interactive AI can mediate coordination and execution, increased scale also risks diluting human interpretive control. This raises questions about how much autonomy can be distributed without eroding the human's role as director, particularly in domains where improvisation and responsiveness are critical.

These challenges are further amplified in emerging creative domains where robots function not only as interactive agents, but also as active materials for design and expression. In contexts involving non-rigid or deformable robotic systems [2], deformable robotic systems [22], or expressive robotic bodies, interaction unfolds through ongoing negotiation rather than discrete instruction. Similar dynamics arise in creative practices such as robotic dance, performance, or musical improvisation [12]; for example, when a robot participates in piano improvisation alongside a human performer (similar to that in [6] (Figure 1 Frame 2)). Interactive AI must therefore scaffold exploration and interpretation without prematurely stabilising behaviour or constraining expressive possibility.

While we ground our discussion in creative contexts, the scaffolding lens extends to any domain where intent evolves over time. Mission-critical contexts introduce additional tension between direction and safety. In time-pressured situations such as disaster response or emergency coordination, humans may need to improvise or "wing it" under uncertainty [14]. Interactive AI can scaffold rapid interpretation and execution, but over-directing or delaying action risks undermining human intuition and accountability (Figure 1 Frame 4). Conversely, insufficient mediation may expose users to safety or trust failures [9]. Designing systems that support human-directed improvisation, so they can "wing-it like McGyver" during such scenarios, while maintaining transparency, trust, and responsibility remains a central challenge.

Taken together, if humans are to remain executive directors rather than supervisors of autonomous systems, then interactive AI must be designed to support interpretation, intervention, and accountability in these systems. Addressing these challenges will require rethinking not only system capabilities, but also how agency and responsibility are made visible and negotiable in human–AI–robot collaboration.

4 Conclusion

As interactive AI and robotics mature, the central question for human–robot interaction is no longer how to maximise autonomy, but how to design systems that know when not to take over. In creative, learning-oriented, and time-critical contexts, effective collaboration is less dependent on optimisation than on the ability to improvise, reinterpret, and adapt as situations unfold. We have argued for re-framing interactive AI as scaffolding: infrastructure that enables humans to shape robotic behaviour while remaining meaningfully in control. By positioning humans as executive directors and AI as mediating partners, this perspective foregrounds interpretation, intervention, and accountability as first-class interactional concerns. Rather than prescribing a specific architecture, we offer a conceptual lens for reasoning about initiative and control in human–AI–robot collaboration. We invite the community to rethink automation-centric narratives and explore how interactive AI can function as enduring scaffolding for human creativity, learning, and responsible action with robots.

References

[1] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.

[2] Alberto Boem and Giovanni Maria Troiano. 2019. Non-rigid HCI: A review of deformable interfaces and input. In *Proceedings of the 2019 on Designing Interactive*

Systems Conference. 885–906.

[3] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.

[4] Mihaly Csikszentmihalyi and Mihaly Csikzentmihaly. 1990. *Flow: The psychology of optimal experience*. Vol. 1990. Harper & Row New York.

[5] Nicholas Davis. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 9. 9–12.

[6] Jordan Aiko Deja, Sven Mayer, Klen Čopič Pucihar, and Matjaž Kljun. 2022. The Vision of a Human-Centered Piano. *arXiv preprint arXiv:2204.06945* (2022).

[7] Paul Formosa. 2021. Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds and Machines* 31, 4 (Oct. 2021), 595–616. doi:10.1007/s11023-021-09579-2

[8] Philipp Graf, Christian Sønderskov Zarp-Falden, Lakshadeep Naik, Kevin Bruno Lefeuvre, Emanuela Marchetti, Eva Hornecker, Mads Bergholdt Sørensen, Laurits Valberg Hemmingsen, Ebbe Vincent Just Christensen, Leon Bodenhagen, et al. 2024. Distributed agency in HRI—an exploratory study of a narrative robot design. *Frontiers in Robotics and AI* 11 (2024), 1253466.

[9] Maciej Grzeszczuk, Grzegorz Pochwatko, Barbara Karpowicz, Stanisław Knapiński, and Wiesław Kopeć. 2025. Building Trustworthy Cognitive Monitoring for Safety-Critical Human Tasks: A Phased Methodological Approach. *arXiv preprint arXiv:2506.22066* (2025).

[10] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. 2019. Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. *Advanced Robotics* 33, 15-16 (2019), 764–799.

[11] Guy Hoffman. 2019. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218.

[12] Guy Hoffman and Gil Weinberg. 2010. Gesture-based human-robot jazz improvisation. In *2010 IEEE international conference on robotics and automation*. IEEE, 582–587.

[13] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.

[14] Hans-Christian Jetter, Leon Raule, Jens Gerken, and Sören Pirk. 2025. Mixed-Initiative Dynamic Autonomy Through Variable Levels of Immersion and Control (MIDA-VIC): A New Paradigm for Collaborative Robotic Teleoperation in Space Exploration. In *Advancing Human-Computer Interaction for Space Exploration (SpaceCHI 2025)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 22–1.

[15] Matthew Jeung, Anup Sathya, Wanli Qian, Steven Arellano, Luke Jimenez, and Ken Nakagaki. 2025. Shape n'Swarm: Hands-on, Shape-aware Generative Authoring with Swarm UI and LLMs. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[16] Philippa Madill, Matthew Newton, Huanjun Zhao, Yichen Lian, Zachary McKendrick, Patrick Finn, Aditya Shekhar Nittala, and Ehud Sharlin. 2025. Playing with Robots: Performing Arts Techniques for Designing and Understanding Robot Group Movement. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[17] Zachary McKendrick, Ori Fartook, Patrick Finn, Ehud Sharlin, and Jessica Cauchard. 2023. Waiting in the wings: Drones in live performance. In *Graphics Interface 2023-second deadline*.

[18] Christina Pasalidou, Chris Lytridis, Avgoustos Tsinakos, and Nikolaos Fachantidis. 2025. Augmented Reality and Robotics in Education: A Literature Review. *Computers in Human Behavior: Artificial Humans* (2025), 100157.

[19] Nicole Robinson, Brendan Tidd, Dylan Campbell, Dana Kulić, and Peter Corke. 2023. Robotic vision for human-robot interaction and collaboration: A survey and systematic review. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–66.

[20] Taishi Sawabe, Suguru Honda, Wataru Sato, Tomoki Ishikura, Masayuki Kanbara, Sakiko Yoshikawa, Yuichiro Fujimoto, and Hirokazu Kato. 2022. Robot touch with speech boosts positive emotions. *Scientific Reports* 12, 1 (2022), 6884.

[21] Tyrone Justin Sta. Maria, Faith Griffin, and Jordan Aiko Deja. 2025. Set the Stage: Enabling Storytelling with Multiple Robots through Roleplaying Metaphors. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[22] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–33.

[23] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.

[24] Ruhan Wang, Shuowen Li, Peiran Zhang, Danqi Huang, Yijie Guo, and Haipeng Mi. 2025. PuppetLine: An Interactive System for Embodied Storytelling with LLM-driven Swarm Robots. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[25] Weitian Wang, Rui Li, Yi Chen, Z Max Diekel, and Yunyi Jia. 2018. Facilitating human–robot collaborative tasks by teaching-learning-collaboration from human demonstrations. *IEEE Transactions on Automation Science and Engineering* 16, 2 (2018), 640–653.