



# DeepCraft: Co-Intelligent Architecture and Human and AI-Driven Craftsmanship in Design-to-Production Pipelines

Peter Buš<sup>(✉)</sup>

Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School,  
University Town of Shenzhen, Nanshan District, Shenzhen 518055, People's Republic of China  
[peter\\_bus@sz.tsinghua.edu.cn](mailto:peter_bus@sz.tsinghua.edu.cn)

**Abstract.** The working paper investigates the potential of artificial intelligence technologies (AI), namely the Generative Adversarial Imitation Learning (GAIL) implemented in a process of digital robotic fabrication prospectively to be used in craftsmanship. The method introduced is based on a preliminary demonstration provided digitally in an abstract toolpath generated by a human-driven movement in a hand gesture translated into a digital space in a real-time process. The investigation presented in this paper focuses on a preliminary computational digital framework which may serve as a base for further investigation. At this stage of the report, the framework encompasses human hand recognition creating a toolpath for a robot, which learns its principles and tries to interpret the process in a digital space. This learned toolpath resulted in a digital brain being applied again in a different shape of the human-created toolpath or gesture movement. The paper also presents the computational system of the real-time navigation of the robot based on a human gesture in a virtual space. The learned knowledge by a robot is observed in a digital environment before any physical applications.

**Keywords:** Intelligent design-to-production · Deep learning in craftsmanship · AI-driven craft · Gesture-driven manufacturing · GAIL

## 1 Introduction

At present, the architecture, engineering, construction and operation sector (AECO) addresses a variety of challenges related to its continuous digitization, low production-efficiency in construction, sustainability, circularity, eliminating carbon emissions, and mass-customization also through the lens of Artificial Intelligence (AI) as a part of revolutionizing Industry 4.0.

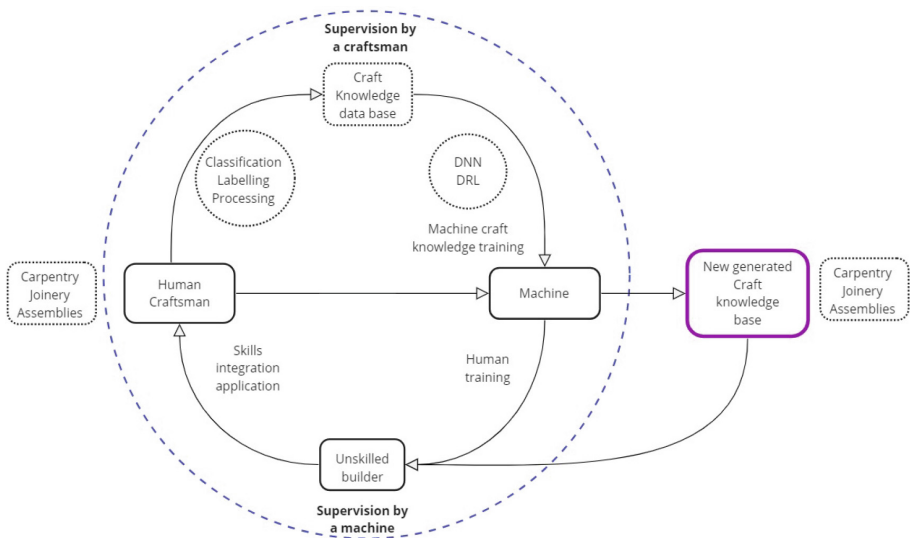
Besides numerically controlled digital fabrication (NC), Building Information Modelling (BIM), design for manufacture, assembly and disassembly method of production (DfMA), and standardization of building components to create architectural scenarios (Kit-of-Parts), the AI field is starting to be prevalent to augment and extend human capabilities to deal with complex and high dimensional problems and requirements within the AECO sector too. The current advancements in these technologies place higher demands

© The Author(s) 2024

C. Yan et al. (Eds.): CDRF 2023, *Phyigital Intelligence*, pp. 368–378, 2024.

[https://doi.org/10.1007/978-981-99-8405-3\\_31](https://doi.org/10.1007/978-981-99-8405-3_31)

on the skills of human designers, production technologists, and construction experts to solve specific problems in mutual human-machine collaborations (Duan et al. 2017). The digitally unskilled workers, although not familiar with the latest advancements in digital fabrication, are usually highly skilled craftsmen (e.g. carpenters in timber construction production) capable of delivering unique and customized products, while preserving the traditional notion of crafts with specific human-made qualities. Thus, the craftsmen bring a qualitative value into the production pipelines, meeting the high standards as well as many qualitative criteria in hand-made processes (aesthetic qualities, detailing, smartness, and complexity of the chosen solution made by a human, hand-made quality, individually recognized authorship, specific artistic and artisanal qualities) (Pu 2020). But how to recognize those qualities in the era of technological and machine intelligence advancements?



**Fig. 1.** Diagram of the human-in-the-loop process engaging with AI technology

## 2 Artificial Intelligence in the Scope of a Craftsmanship

This paper theoretically envisions that through human and artificial intelligence-driven processes of digital fabrication and production of artifacts where the craft skills are recognized, learned, trained, and implemented in the human-in-the-loop co-creative production workflow such as in the one-shot imitation learning (Finn et al. 2017), the technology will be capable to develop and strengthen its “wisdom” in a similar way how humans improve their skills, experience, and wisdom in time and thus, make autonomous decisions in the production process. Can a machine be capable of conceptualizing learned knowledge to yield novel artifacts through hybridized/synthesized modes of human-machine interactions utilizing Neural Networks (NN) and deep reinforced learning (DRL)?

There is a hypothesis suggesting that unique knowledge demonstrated by craftsmen can be translated into forms of demonstrations, digital models, additional data sources (such as images, videos, and sequences), and digital processes (generation of tool paths for specific digital robotic fabrication execution, unique assembly modes).

The paper further proposes that by linking human intelligence (knowledge, experiences, craft skills, capacity to make relevant decisions) and machine intelligence (responsive robotics based on multisensorial setup (Felbrich et al. 2022), XR devices and digital operations) in one coherent hybridized production loop, there will be created a novel communication and interaction platform between human and a machine via physical interventions and demonstrations leading to machine capabilities improvements to execute the production task.

At this stage, the paper introduces a computational connection between a human agent and an AI agent in a digital process to create an abstract toolpath, while theoretically envisaging a novel searching and generation method for design and production space based on a human-in-the-loop cooperative learning process (Fig. 1). In this report, the human and machine agents consider human-driven toolpath generation in a form of an intuitive gesture, while human logic, preferences, memory, cognition, and past experiences will be investigated in the next phase of the research. These aspects are tested to be transferable to a machine in a continuous training and learning process driven by a human. Consequently, the machine will constantly improve its capability based on human agents' inputs and become more autonomous in the decision-making and generation of design and production space. Such a framework aims to imitate in a re-interpreted event a human gesture, e. g. for drawing or painting intervention in a digital space.

Instead of replication and recreation of the crafting process in a numerically controlled way of digital fabrication (NC), the intention is to discover a way how the machine can express itself in a novel and augmented craft language and its formal expression in artefacts as similar as a human does, but unconventionally, beyond human imaginative solution craft space. The paper provides a preliminary conceptual digital experiment using the digital twin of the desktop robotic arm equipped with a virtual multisensorial setup. The machine learns a simple human-driven toolpath, considering gentle human movements of the hand physically provided by the demonstrator, translated into the digital space.

## **2.1 Current Learning Experiments in the Processes of AI in the AECO and the Use of GAIL**

AI deep learning implementations in the processes of robotic digital fabrication for the potential use in the AECO sector have been tested in a variety of tasks, such as the assembly of a lap joint (Apolinarska et al. 2021) or pick and place scenarios for component assemblies (Felbrich et al. 2022). Other studies focus on co-designing strategies for autonomous construction methods (Menges and Wortmann 2022), exploring the integration of deep reinforcement learning for the intelligent behaviour of construction robots as builders.

The question of how to involve human agency in the AI-driven processes to achieve coherent results for the potential use of AI in the AECO applications on a larger scale

or in human-made operations is still broadly unexplored. Imitation learning, especially Generative Adversarial Imitation Learning—GAIL (Ho and Ermon 2016) as a method to teach a robot to do a task has a solid potential to be integrated into the design-to-production processes if we consider smaller scale in the early stage of production, such as drawing or cutting toolpath. At present, pick-and-place scenarios of simple objects using the visual demonstration and data collected from a human agent are successfully deployed (Finn et al. 2017) as a combination of the imitation and meta-learning strategies, however, the movements of the robot are still very technical and pre-programmed, although delivering the simple task successfully. The Unity and ML Agents tool to train a robot have been previously introduced by Pinochet as Smart Collaborative Agents (2020) and Hahm (2020), when the robot follows the pre-defined targets, not using imitation learning approach, and based on configurable joints or Unity articulation body.

The notion of human craftsmanship with a unique look is a very complex task to imitate or deliver, as it requires a complexity of information and data to be collected and processed. This research explores how to engage the human agent with a robot, aiming to find a method for a process to either participate together in a real-time sequence scenario or the human can act as an expert and demonstrator to teach the robot a task to execute.

### 3 Implementing the Computational Framework

#### 3.1 Real-Time Gesture-Driven Navigation of the Robot

There are two computational implementations described in this paper, integrating the real-time robotic navigation by a human gesture: Unity and Rhino/Grasshopper environments, which serve as an initial input for the gesture-driven toolpath generation. Both are described in detail in the following sections (Figs. 2 and 3).

The real-time navigation of a robot may have a variety of possible uses in the field of digital fabrication and manufacturing for architecture. The proposed framework can also serve as a designer's environment to be tested and explored before any manufacturing and crafting process. The data captured from a human can be stored and implemented in a custom scenario. Even though the implementation at this stage is not fully practical due to specific constraints related to noisy interference of the data exchange, the real-time interaction is engaging and can serve for further investigation in connection with the real robot. The computational models for both strategies are available from Buš (2023) and GitHub (2023).<sup>1</sup>

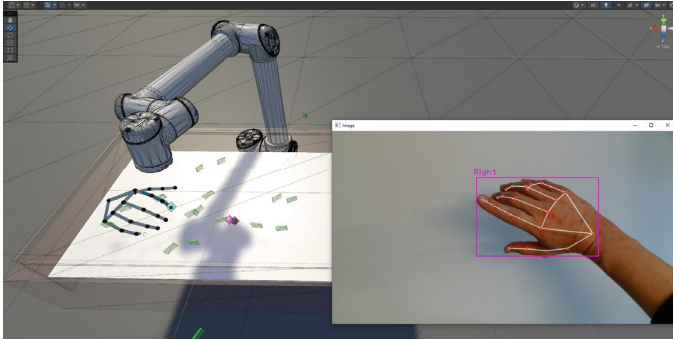
#### 3.2 Unity and Rhino/GH Implementation—Hand Tracking

Both implementations encompass the User Datagram Protocol library (UDP) for the data transfer between the actual gesture and the digital environment and the robotic twin. For this implementation, the Universal Robot UR1 digital twin and a standard web camera have been used for human hand capturing.

---

<sup>1</sup> Buš, P. Repositories [Internet][Shenzhen], Github; 2023 [updated 2023 May 4, cited 2023 May 5]. Available from <https://github.com/peterbus?tab=repositories>.

The computational approach of the hand recognition and the data exchange platform in Unity implement the Unity UDP receiver script, provided by the CV Zone platform as an open data resource (Murtaza 2022). The Unity and RhinoGrasshopper environments were customized and adapted for the robotic movement. Both strategies utilize the CVZone Hand Tracking module with the Hand detector implemented in Python to recognize the hand of a human (Murtaza 2022). Hand recognition contains 21 points which are interconnected with lines, representing the virtual skeleton of the human hand.



**Fig. 2.** Hand tracking implementation in Unity connected to the UR robot. The robot follows the finger in real-time process while properly rotating based on configurable joints

### 3.3 Unity Basic Setup

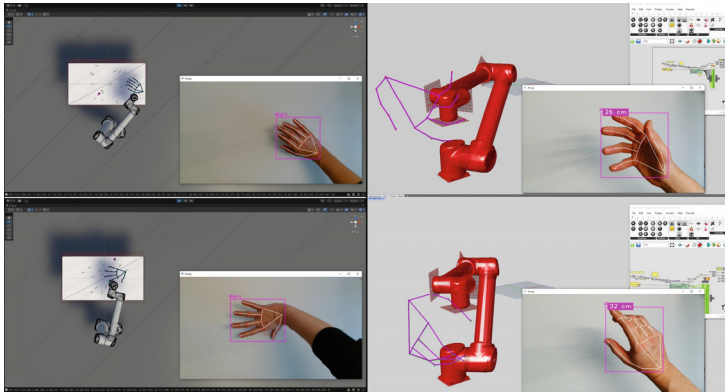
The 21 recognized points are transferred utilizing the UDP data protocol into the Unity environment using the local host. The data are constantly received by the receiver and the points are embedded as Game Objects creating the foundation for the skeleton.

The specific point can be selected as a spawner of the checkpoints for the robotic toolpath. Based on human movement, the hand spawns the targets for the robot, specifically rotated according to the hand movement. Custom C# scripts were written to link the hand with the digital model of the UR robot, which is based on configurable joints for each of its axes. As such, it was possible to create a Target for the robot, which follows it. In that way, the robot is navigated by the hand point on the selected finger in real time, considering the physics engine in Unity and rotating and moving based on the customized configurable joints.

### 3.4 RhinoGrasshopper Setup

Similarly, the recognized hand points were transferred via the UDP protocol into the RhinoGrasshopper environment and points were reconnected. This was done as an independent platform. For the UDP communication transfer, the GHowl addon has been used (Alomar et al. 2011), considering the position of the points as well as the distance information between the hand and the web camera.

Using this information, it was possible to implement the third dimension to navigate a virtual end effector of the robot in all three dimensions. For the robotic real-time simulation of the moving robot, the Robot addon was utilized. The GH definition can serve as a starting point and a test bed for further implementations and testing purposes. The working version of the GH definition is available<sup>2</sup> (Buš 2023; GitHub 2023).



**Fig. 3.** Testing the robot and gesture movements in real-time. The hand spawns the targets for the robotic toolpath, according to the direction of the hand movements. The framework was implemented in Unity and Rhino Grasshopper utilizing UDP, Ghoul and Robots addons

### 3.5 GAIL and Behavioral Cloning Test in Unity and Observation

The Unity environment was further tested to teach the robot to recognize the human gesture and interpret it afterward it was captured. Several custom scripts were developed to do so as well as the standard toolpath following a system based on numerically controlled positions. These were captured from the spawned checkpoints from the human gesture to create a linear toolpath. However, the implementation of the deep learning method, in this case, imitation learning utilizing the ML-Agents tool in Unity with the GAIL learning method combined with behavioral cloning was tested and observed (Juliani et al. 2018).

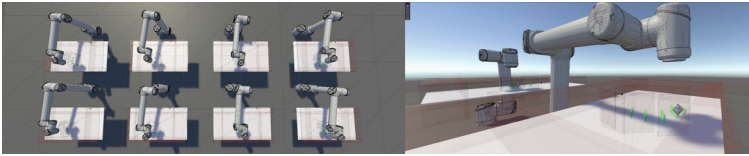
The GAIL (Ho and Ermon 2016) considers the policy from the expert demonstration to perform a task based on ‘how to act by directly learning a policy’ from the data provided. The ML-Agents tool contains an imitation learning approach utilizing the GAIL and the behavioral cloning method, which aims to capture the pre-defined process of a demonstration of how the robot should perform the task according to the expert demonstration. It follows the sequence of targets in a toolpath, previously generated by human in real time. In this experiment, the data captured from the gesture served as

<sup>2</sup> Buš, P. Hand-Tracking-to-navigate-a-robot-Rhino-Grasshopper-framework.[Internet][Shenzhen], Github; 2023 [updated 2023 May 4, cited 2023 May 5]. Available from <https://github.com/peterbus/Hand-Tracking-to-navigate-a-robot--Rhino-Grasshopper-framework>.

an input for the demonstration recording, containing the transform position information (transform position, rotation, scale) of spawned targets from the gesture.

The positions were translated into the toolpath, and a virtual ML agent run through them several times (see Fig. 4). The agent can serve later as an input for the robotic end effector target, mentioned above. The heuristics training simulation contains the digital demonstration, captured as a demo for the GAIL and behavioral cloning training.

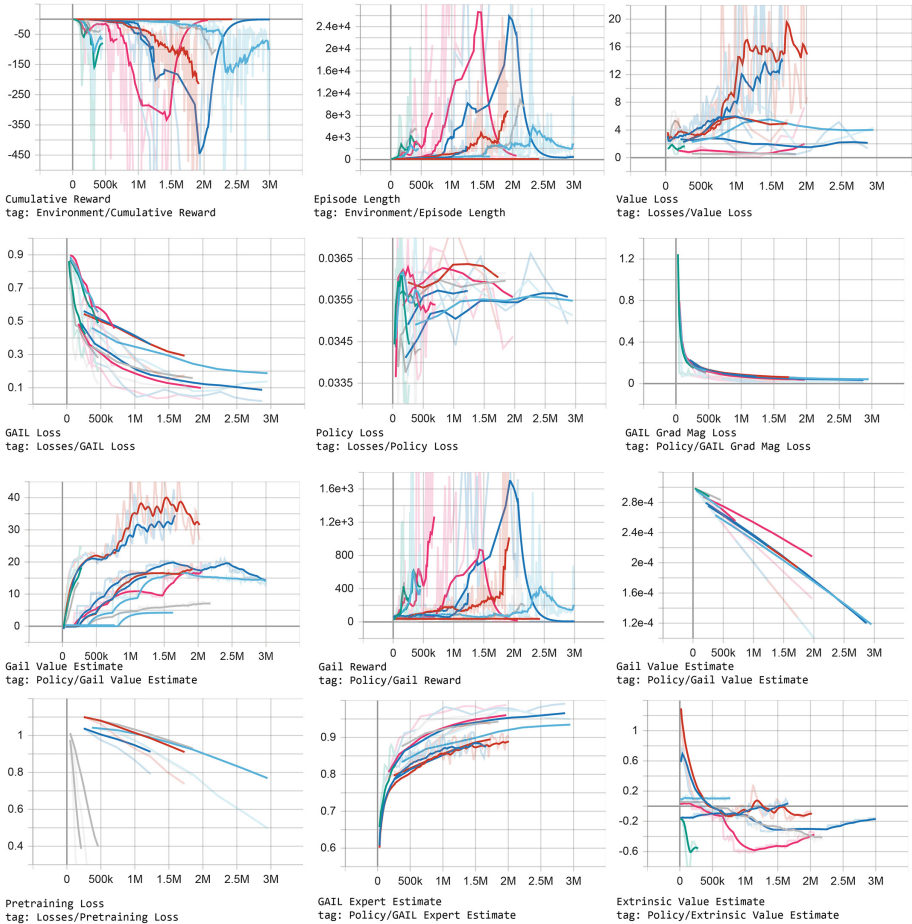
The training algorithm included the default one based on the Proximal policy optimization hyperparameters (PPO) for the ML agent, tested with different setups for the GAIL strength or behavioral cloning. The virtual agent looks randomly for the checkpoint positions in space and learns from them how to interact with them in each of the episodes. The task for the agent was to recognize starting position and end position as well as checkpoints to perform the toolpath in the right order and direction. In addition, each of the iterations slightly randomly moves the positions of the path checkpoints to encourage the agent to learn from these novel positions. This might serve for the potential future gestures that will be each time different.



**Fig. 4.** The virtual end effector follows the agent, while running through the checkpoints on the toolpath

The learning process contained 3–5 million iterations (steps) with a positive or negative reward structure for the agent, each time it collides with the correct or wrong checkpoint. The process generated the virtual brain for future testing scenarios. As it was qualitatively observed from the preliminary tests in the learned positions of the agent in the final inference training, the results with the current setups do not precisely imitate the original demonstration, although the agent reaches the targets in the right directions and with the right orientation in a sequential way. The quantitative results are provided in the following scalars, captured from the Tensorboard platform (TensorFlow 2023), showing relevant reward processes and GAIL policies. While the cumulative reward is decreasing (in case the model was trained without extrinsic rewards), the GAIL Loss and Pretraining Loss showed some models adopted well according to the demonstration, as the curve slightly decreases in time, assuming the agent learns the policy. The GAIL reward increased after a certain number of iterations and the agent obtained relevant rewards while learning the policy. There was a big decrease observed at the beginning of the training process—it depends on a variety of combinations of hyperparameters set in the configuration file. The training delivers a variety of brains with less or acceptable training results. During the training, each scenario had a moment of decreasing the reward value, which later became more stabilized. In addition, the agent continuously improves the imitation of the demonstration during the training duration (Fig. 5).





**Fig. 5.** Preliminary GAIL training results of the agent in several running scenarios. Its scalars show the reward and episode length increase, although there is a big drop in the first part of the training process. The Policy loss is not convincing at the preliminary stage for some cases-it should rather decrease. Additional set of observations related to GAIL policies

## 4 Discussion and Further Potential

Although the preliminary training of the agent to follow the toolpath is not satisfactory enough, the Unity-based computational framework might serve as a base for further testing and observations. So far only one type of algorithm has been tested, namely the Proximal Policy Optimization (PPO), which uses a neural network to approximate the ideal function that maps an agent's observations to the best action an agent can take in a given state (Juliani et al. 2018).

The other algorithms and different hyperparameters can be tested and evaluated according to the specific needs of the designer, such as testing different strengths of GAIL or behavioral cloning and their combinations. The potential of the human hand,



movements, and gesture recognition lies in the prospective implementation in the making and crafting processes when the hand and movements of the craftsman can be captured and recognized to inform the learning policy in a form of an expert demonstration.

At this stage of the investigation, the robot movement is not straightforward, as it contains a certain noise, which prevents the robot from moving smoothly as in the demonstration. This can be addressed by a higher number of training episodes during the default training (this also requires longer training time), and a higher number of steps in the demonstration data. The robot itself can be set up through an updated articulation body tool in Unity, benefiting from Unity physics instead of the current setup of configurable joints. This will improve the motion of the robot. Even though the gesture is not precisely cloned by AI, the resulting digital process partially follows the human inputs because of a pre-trained process. In the next phase of the research, such an approach can be used to train the AI in the assembly of spatial scenarios based on components deployed as kit-of-parts to create a spatial configuration. In the context of the AECO, this might contribute to the space created in unconstrained construction site conditions, while considering the human aspect in creating a unique space.

The future research will concentrate also on the demonstrations provided by the craftsmen, utilizing more advanced recognition-based sensorial setups, such as motion capture methods and tactile sensors to obtain more precise data. These will be integrated with the Unity framework. From the preliminary results, the author observed that the GAIL combined with behavioral cloning (the strength ratio implemented was 0.5 for both reward signals-GAIL and behavioral cloning) has the potential in digital fabrication and production processes, however, more tasks and more robust processes must be tested first, such as the creation of an assembly based on kit-of-parts system.

## 5 Conclusion

This working paper introduced preliminary computational frameworks for further testing and observations, potentially to be deployed in handcrafting or assembly processes, utilizing digital tools, such as collaborative robots.

The environments such as Unity and Rhinoceros can serve as platforms to integrate more gentle operations in making, based on handcrafting, followed and learned by AI. Even though this hypothesis has not been fully proved yet as the computational models need further development and testing, it may be argued that hands-on operations followed by AI-driven technologies will shift the way how the crafting processes can be executed in the future and will bring novel understanding where the human agent is still an expert and an important production agency in the human-in-the-loop processes.

Preliminary observations of the virtual hand proved satisfactory real-time navigation of the robot (without a specific sensorial framework), however, further testing with the physical robot is necessary to fully prove the concept.

## References

- Alomar, D., Fraguada, L.E., Piacentino, G.: Ghowl Addon (2011). Food4Rhino <https://www.food4rhino.com/en/app/ghowl>
- Andrychowicz, M., Stadie, B., Jonathan, H.O., et al.: One-shot imitation learning. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 30, pp 1087–1098. Curran Associates, Inc. (2017). <http://papers.nips.cc/paper/6709-one-shot-imitation-learning.pdf>
- Apolinarska, A.A., Pacher, M., Li, H., et al.: Robotic assembly of timber joints using reinforcement learning. *Autom. Constr.* **125**, 103569 (2021). <https://doi.org/10.1016/j.autcon.2021.103569>
- Buš, P.: Repositories [Internet][Shenzhen], Github; 2023 [updated 2023 May 4, cited 2023 May 5]. Available from <https://github.com/peterbus?tab=repositories>
- Duan, Y., Andrychowicz, M., Stadie, B., Jonathan, H.O., et al.: One-shot imitation learning. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 30, pp 1087–1098. Curran Associates, Inc. (2017). <https://papers.nips.cc/paper/6709-one-shot-imitation-learning.pdf>
- Felbrich, B., Schork, T., Menges, A.: Autonomous robotic additive manufacturing through distributed model-free deep reinforcement learning in computational design environments. *Constr. Robot.* **6**, 15–37 (2022). <https://doi.org/10.1007/s41693-022-00069-0>
- Finn, C., Yu, T., Zhang, T., Abbeel, P., Levine, S.: One-Shot Visual Imitation Learning Via Meta-Learning (2017). Available at: <http://arxiv.org/pdf/1709.04905v1>
- Github: Let's build from here (2023). [Internet][Place unknown], Github; 2023 [updated 2023, cited 2023 May 5]. Available from <https://github.com/>
- Hahm, S.: Training robot arm with Unity ML agents (2020). [Internet][Place unknown], Youtube; 2020 [updated 2020; cited 2023 May 5]. Available from <https://www.youtube.com/watch?v=HOUPkBF-yv0>
- Ho, J., Ermon, S.: Generative Adversarial Imitation Learning (2016). <https://doi.org/10.48550/arXiv.1606.03476>
- Juliani, A., Berges, V.P., Vckay, E., et al.: Unity: A General Platform for Intelligent Agents (2018). ArXiv:abs/1809.02627
- Menges, A., Wortmann, T.: Synthesising Artificial Intelligence and Physical Performance, in Machine Hallucinations Architecture and Artificial Intelligence. AD, Wiley (2022)
- Murtaza, H.: CVZone (2022). <https://www.computervision.zone/>
- Pinochet, D.: Digital Futures 2020—Smart collaborative agents (2020). [Internet] [Place unknown], Youtube; 2020 [updated 2020; cited 2023 May 5]. Available from <https://www.youtube.com/watch?v=KDObBwoyzKg&t=771s>
- Pu, J.: Integration of arts and crafts in artificial intelligence environment. *J. Phys.: Conf. Ser.* **1574**, 012162 (2020). <https://doi.org/10.1088/1742-6596/1574/1/012162>
- TensorFlow. TensorBoard: TensorFlow's visualization toolkit (2023). [Internet][Place unknown], Google Brain team; 2023 [updated 2023, cited 2023 May 5]. Available from <https://www.tensorflow.org/tensorboard>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

