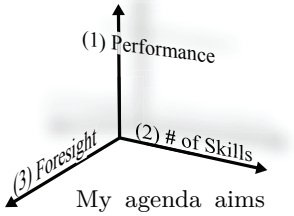
A Scalable Recipe for Robot Learning from Synthetic Data

Ge Yang

Despite decades of progress, robots remain confined to controlled lab spaces and factory floors, performing a few specialized tasks. They move conservatively, struggle with handling rich contacts, and perform poorly when it comes to making long-term decisions in open, uncertain settings.

My research aims to give robots **human-like agility and dexterity** for physical interactions, along with the **intelligence** to solve **complex, long-horizon tasks**. Central to this mission is finding ways to scale up *learning* and *data*. While many [1, 2, 3] advocate collecting more real-world demonstrations to achieve generalist robots, I focus instead on developing **AI-powered virtual learning environments** that leverage generative models to **create data**. Through generative AI, we can make synthetic data significantly more abundant and systematically varied than real-world datasets—bringing us closer to solving key challenges in both robotics and artificial intelligence.

My past and ongoing research encompass three main thrusts: (1) **mastering** real-world visuomotor control via synthetic data, (2) **scaling skill-wise** towards a general-purpose robot foundation model, and (3) learning to achieve complex tasks that require **foresight**. **The central theme is leveraging virtual environments to scale up robot learning.** Thrust 1 establishes its feasibility; Thrust 2 expands the number of applicable skills; Thrust 3 extends the temporal context length to tackle much more difficult tasks.



My agenda aims to expand the frontier along three critical axes.

Thrust 1: Can A Robot Learn from Machine Dreams?

AI breakthroughs in the past decade were mainly driven by increasing the quantity and quality of data. Unlike images and text that are abundantly available from online sources, robotic datasets are magnitudes more scarce. For instance, the LION 2B dataset [4] used to train the text-to-image generative model, StableDiffusion [5], contained 2.3 billion images. This makes even the largest robot dataset today [3] 10^5 times too small based on the generous assumption that $10k$ hours of robot data contain $30k$ distinct scenes. In reality, robotic data is much less diverse.

My work, **LucidSim** [1], established a promising new direction for closing this data gap. LucidSim is an AI-powered physics simulator that enables training **real-world visual policies without real-world data**. Vision researchers have long dreamed about this, but getting it to work requires solving three difficult problems. First, image diffusion models were developed to make beautiful pictures, but

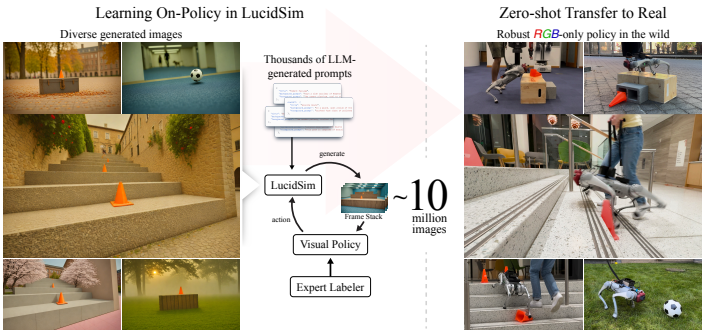


Figure 1: **Learning Visual Parkour from Generated Images.** (left) Using generated images from LucidSim, we can produce (right) robust real-world visual parkour policies.

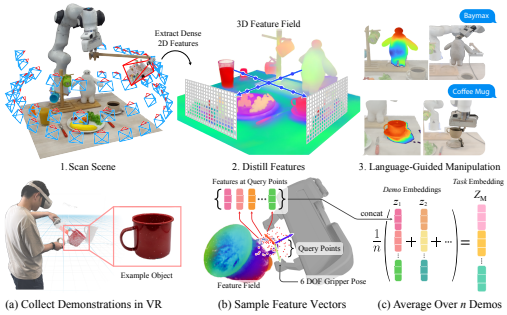


Figure 2: **CLIP Feature Fields for Manipulation** is a higher-order representation of the gripper ↔ scene relationship.

our robots need accurate physical dynamics and geometry. Second, the image generation pipeline is $75 \sim 350\times$ slower than the wall clock. Third, naïvely training on off-policy trajectories performed poorly, even with significantly larger datasets.

LucidSim provided all three solutions at once. First, I made the data realistic and diverse by building around the MuJoCo physics engine and using the object mask and depth render from the simulator to compose and condition the generated images. This ensures visual consistency with the scene geometry. To mitigate the loss of sample diversity due to such conditioning, I injected variations by sourcing structured image prompts from ChatGPT. Second, I improved the rendering speed with a new technique, Dreams In Motion (DIM), which warps a single generated image to consecutive frames via the optical flow computed from changes in the robot's camera pose and the scene geometry. DIM made LucidSim a magnitude faster. Finally, to generate on-policy data, we needed to run image generation in a closed loop where the visual policy takes in generated images at each time step. Getting LucidSim "go burrr" was key, so I developed systems tools to distribute trajectory sampling, image warping, and image generation across 80+ GPUs. LucidSim's results indicated that closed-loop training was solely responsible for getting the visual policy to reach expert-level performance.

LucidSim is the latest milestone in my quest to expand robotics capabilities. In early 2022, my work *rapid locomotion* [2] achieved **record running speeds** on an MIT Mini Cheetah. In 2023, neural volumetric memory **integrated depth-perception** to enable autonomous stairs-climbing [3]; continuing into 2024, my UC San Diego collaboration produced the first **learned humanoid whole-body control** [4]. This was followed a few months later by cross-embodied dexterous object manipulation on **multiple humanoid platforms** [5]. The common theme is using closed-loop training in simulation to attain the vital marker of **a scalable learning system**—a monotonic curve that translates additional compute into policy improvement (see Fig.12 in the paper [1]). This series of results positions LucidSim as a promising data source for robotics' next hundred trillion tokens.

Future Work 1.1. Human-level athleticism and expressivity. Humanoid and dexterous hand hardware are improving rapidly as we speak, which opens up a rich space for expressive, personable, and athletic robots that were less than accessible in the past. Substantial research is needed, and my plan is to study competitive sports (think tennis), load-bearing, whole-body manipulation (think of shouldering a large box or elbowing through a fire door), and finally, moving with personality and intent. I want robots that are not only capable but can also bring us joy and enrich our lives.

Thrust 2: Towards A Generalist Robot Foundation Model

This thrust aims to expand visuomotor learning skill-wise and build a generalist robot that can generalize to new skills with little additional training data. Before 2022, I put substantial effort into understanding how neural networks generalize during a decision process [6, 7, 8]. Most notably, I made significant progress on a textbook problem in deep RL called the "deadly triad" [6] by explicitly considering learning as a dynamical process approximating kernel regression [7] as opposed to viewing neural networks as static, function approximators defined by an "expressivity."

In late 2022, it became clear to me that the knowledge to generalize was not in clever algorithms but in the data, and training from scratch on small, bespoke datasets for a task will not be how we build robots in the future. I began considering ways to represent human demonstrations with features from vision and vision-language foundation models (VLMs). This view differs from prior work representing scenes or objects by formulating the problem as **learning a higher-order representation of the relationship** between the robot gripper and its surrounding context. In the resulting

method F3RM ([9], see Fig.2), I "baked" 2D image features from multiple camera perspectives (from CLIP [7]) into a 3D neural radiance field [8, NeRF], bring example grasps into a shared semantic space with natural language.

F3RM demonstrated **strong open-set generalization** to unseen text queries and entirely new categories of objects while using only five example objects and twelve grasp examples (in total). This work was selected to be **The Best Paper** (top 0.2%) out of 499 submissions at the 7th Conference of Robot Learning (2023). It inspired follow-ups that explore higher-order scene formats [9]; extension to 3D Gaussian Splatting [10]; and manipulation with a mobile robot [10, 11, 12] [11].

Language prompts and human demonstrations are natural ways to specify a manipulation task, but being tethered to a physical robot setup limits how quickly we can scale to diverse scenarios and tasks. So, I began developing mixed reality (XR) tools that have since become a go-to solution for imitation learning manipulation tasks on humanoids [5]. Below, I will outline my approach to overcoming key challenges in training a generalist robot foundation model from synthetic data.

Future Work 2.1. Internet-scale foundry for human demonstrations. I plan to collect an **internet-scale synthetic robot dataset** by crowd-sourcing via the web. I have *two* unique advantages: First, LucidSim enables training real-world robot skills from virtual reality demos that, before LucidSim, were not as useful. Second, I recently moved the physics engine MuJoCo, which typically runs on a separate, tethered computer, to run **natively on the VR device** via the web browser. This was technically very difficult as it involved compiling to WebAssembly and rewriting the entire graphics front-end. My plan is to deliver physically realistic, zero-latency immersive experiences through the browser to crowd-source robot demos. This acts as a great setup for the next agenda item (**F2.2.**).

F2.2. Generating worlds, scenarios, and other actors. Using generative AI to design environments and scenarios has the potential to resolve a key scaling bottleneck and is an essential part of my future agenda. Recent work has explored ways to do so, but few have modeled other agents and actors. In autonomous driving, for instance, Waymo's evaluation pipeline models other cars and pedestrians by playing back recorded trajectories. Learning interactive agent models from passive data is not trivial [13, 14], but it is essential if we want robots to occupy human spaces. I believe actor models of increasing levels of sophistication will play a key role in training and enabling closed-loop evaluation.

F2.3. Scaling up closed-loop training to thousands of tasks. It is difficult to manually specify the reward for thousands of skills. However, on the systems level, this **learning pipeline is automated**, which makes it possible to experiment faster at scale than human-in-the-loop setups. This will accelerate our development of AI reward designers and environment builders. On the learning side, recent work mapped RL to a reward regression problem (AWR 15, 16), where the policy is learned entirely by sequence modeling on trajectories and sparse reward [17, 18]. There is a large space for method and empirical innovations that I am confident will enable us to scale.

Thrust 3: Learning to Achieve Long-Horizon, Complex Tasks

An agent in a Markov decision process needs to connect its current decisions with their future outcomes. This becomes significantly harder when the goal is only reachable via a long sequence of actions. During training, a longer horizon makes sampling high-value scenarios less likely; at test time, it demands extended foresight. Many open problems in robotics and AI fit this description. This is evident with chatGPT and self-driving cars—once they became generally capable, our expectations grew to include harder problems that require reasoning and foresight. In the case of chatGPT, this

means solving math problems or writing complex computer programs; in autonomous driving, it means knowing when to safely drive onto the opposing lane to get around a construction site.

This thrust builds upon my prior work on "learning to plan" [12, 13, 14], and aims to give robots and AI the foresight to solve complex, long-horizon tasks. Drawing lessons from successes in computer Go [19], Poker [20], the board game Diplomacy [21], and more recently, Olympia-level mathematics [22], my plan pursuits three angles of attack: the policy, the simulator, and closed-loop evaluation.

F3.1. Making flexible and adaptable policies by planning. Policies used in contemporary deep RL implementations are weak planners with zero look-ahead. This setup can produce highly performant policies, but they tend to collapse to a subset of behaviors and can be inflexible when additional constraints appear at test time. This is a prevalent problem that is especially prominent with humanoids, whose high degrees of freedom induce a space of degenerate solutions, all achieving similar utility. RL only finds some solutions, causing the robot to struggle in cluttered spaces. The goal is to produce flexible and expressive policies that adapt robustly to test-time constraints.

F3.2. Learning to achieve long-horizon, complex tasks. Today, vision language models (VLMs) are not being trained to make decisions. Instead, they are trained to answer single-turn questions for image comprehension. Unlike supervised learning, in a Markov decision process, sampling data involves effort, as the agent has to make a sequence of decisions to reach a state. This is why AlphaGo needed self-play to create its training data. In other words, the data that makes VLMs truly intelligent does not yet exist. How do we generate such data, and in what kind of simulator?

Many researchers today are excited about using video generation models as world simulators for robots. I take a contrarian view and argue that task planning operates at a coarser timescale, where Newton's Laws are less relevant than the chain of events. I propose **"key-frame playground,"** an AI-powered visual text game that extends LucidSim's synthetic data approach to task planning by ignoring short-term physics. I want to build an "AI Dungeon Master" that takes player actions via text at each turn and returns an image. Recent work observed that images interleaved with text produce stronger VLMs [23]. I believe there will be a general method that teaches both robots and AI causal relationships that span many intermediate interactions.

F3.3. Trustworthy and accessible closed-loop evaluation at scale. As robots master diverse and complex skills, they will outgrow existing real-world and simulated evaluation methods. This makes it difficult to measure progress and compare rigorously between labs. **I consider making trustworthy, closed-loop evaluation a centerpiece of my agenda**, and I have been driving a community effort to build such tools in collaboration with colleagues at USC and UC San Diego. LucidSim championed closed-loop evaluation using high-fidelity digital twins in robot parkour, but the methodology is generally applicable to robotics, LLMs, and VLMs, where evaluation is needed for monitoring training collapse. In comparison, LLM and VLM evaluation benchmarks today are mostly limited to a single round of interaction, which is far too simple given what we expect these models to accomplish.

Conclusion

I aim to pave the way toward a new generation of intelligent machines that elevate and amplify our lives. This quest will take most of my research group's effort in the next few years, as there is plenty of space for impactful research in both robot learning and artificial intelligence.

I am confident that this agenda will take us there, and I look forward to supporting colleagues to bring their ideas and algorithmic innovations to life on these powerful platforms.

Publications

[1] **Ge Yang***, Alan Yu*, Ran Choi, Yajvan Ravan, John Leonard, and Phillip Isola. Learning visual parkour from generated images. In *8th Annual Conference on Robot Learning*, 2024.

[2] **Ge Yang***, Gabriel B. Margolis*, Kartik Paigwar, Tao Chen, and Pulkit Agrawal. Rapid locomotion via reinforcement learning. In *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, June 2022.

[3] Ruihan Yang and **Ge Yang** and Xiaolong Wang. Neural Volumetric Memory for Legged Locomotion Control. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2023.

[4] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, **Ge Yang**, and Xiaolong Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[5] Xuxin Cheng, Jialong Li, Shiqi Yang, **Ge Yang**, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. *8th Conference on Robot Learning*, 2024.

[6] **Ge Yang***, Xiang Fu*, Pulkit Agrawal, and Tommi Jaakkola. Learning task informed abstractions. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3480–3491. PMLR, 18–24 Jul 2021.

[7] **Ge Yang***, Anurag Ajay*, and Pulkit Agrawal. Overcoming the spectral bias of neural value approximation. In *International Conference on Learning Representations*, 2022.

[8] **Ge Yang***, Takuma Yoneda*, Matthew R. Walter, and Bradly C. Stadie. Invariance through latent alignment. In *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, June 2022.

[9] **Ge Yang***, William Shen*, Alan Yu, Jansen Wong, Leslie Pack Kaelbling, and Phillip Isola. Distilled feature fields enable few-shot language-guided manipulation. In *7th Annual Conference on Robot Learning*, 2023.

[10] **Ge Yang***, Rizhao Qiu*, Weijia Zeng, and Xiaolong Wang. Language-driven appearance and physics editing via feature splatting. 2023.

[11] Ri-Zhao Qiu, Yafei Hu, **Ge Yang**, Yang Fu, Jianglong Ye, Jiteng Mu, Ruihan Yang, Yuchen Song, Nikolay Atanasov, Sebastian Scherer, and Xiaolong Wang. Learning generalizable feature fields for mobile manipulation. 2023.

[12] **Ge Yang***, Amy Zhang*, Ari S. Morcos, Joelle Pineau, P. Abbeel, and Roberto Calandra. Plan2vec: Unsupervised representation learning by latent plans. In *L4DC*, 2020.

[13] Thanard Kurutach, Aviv Tamar, **Ge Yang**, Stuart J Russell, and Pieter Abbeel. Learning plannable representations with causal infogan. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[14] Lunjun Zhang, **Ge Yang**, and Bradly C Stadie. World model as a graph: Learning latent landmarks for planning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12611–12620. PMLR, 18–24 Jul 2021.

References

[Ref1] Kevin Black. From octo to π0: How to train your generalist robot policy. Talk at X-Embodiment Workshop, Conference on Robot Learning (CoRL), 2024. Workshop presentation.

[Ref2] Open X-Embodiment Collaboration. Open X-embodiment: Robotic learning datasets and RT-X models. October 2023.

[Ref3] Kevin Black and Noah et.al. Brown. π_0: A vision-language-action flow model for general robot control. October 2024.

[Ref4] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *arXiv preprint arXiv:2210.08402*, 2022.

[Ref5] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv [cs.CV]*, December 2021.

[Ref6] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[Ref7] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021.

[Ref8] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[Ref9] Qianxu Wang, Congyue Deng, Tyler Ga Wei Lum, Yuanpei Chen, Yaodong Yang, Jeannette Bohg, Yixin Zhu, and Leonidas Guibas. Neural attention field: Emerging point relevance in 3d scenes for one-shot dexterous grasping. In *8th Annual Conference on Robot Learning*, 2024.

[Ref10] Yuhang Zheng, Xiangyu Chen, Yupeng Zheng, Songen Gu, Runyi Yang, Bu Jin, Pengfei Li, Chengliang Zhong, Zengmao Wang, Lina Liu, Chao Yang, Dawei Wang, Zhen Chen, Xiaoxiao Long, and Meiqing Wang. Gaussiangrasper: 3d language gaussian splatting for open-vocabulary robotic grasping. *IEEE Robotics and Automation Letters*, 9:7827–7834, 2024.

[Ref11] Mingxi Jia, Hao zhe Huang, Zhewen Zhang, Chenghao Wang, Linfeng Zhao, Dian Wang, Jason Xinyu Liu, Robin Walters, Robert Platt, and Stefanie Tellex. Open-vocabulary pick and place via patch-level semantic maps. *ArXiv*, abs/2406.15677, 2024.

[Ref12] Georgios Tziafas, Yucheng Xu, Zhibin Li, and Hamidreza Kasaei. 3d feature distillation with object-centric priors. *ArXiv*, abs/2406.18742, 2024.

[Ref13] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. Human motion trajectory prediction: a survey. *Int. J. Rob. Res.*, 39(8):895–935, July 2020.

[Ref14] Zhiyu Huang, Xinshuo Weng, Maximilian Igl, Yuxiao Chen, Yulong Cao, Boris Ivanovic, Marco Pavone, and Chen Lv. Gen-drive: Enhancing diffusion generative driving policies with reward modeling and reinforcement learning fine-tuning. *arXiv preprint arXiv:2410.05582*, 2024.

[Ref15] Xue Bin Peng, Aviral Kumar, Grace Zhang, and Sergey Levine. Advantage-Weighted Regression: Simple and Scalable Off-Policy Reinforcement Learning. *arXiv [cs.LG]*, October 2019.

[Ref16] Piotr Kozakowski, Lukasz Kaiser, Henryk Michalewski, Afroz Mohiuddin, and Katarzyna Kanska. Q-value weighted regression: Reinforcement learning with limited data. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, July 2022.

[Ref17] Ilija Radosavovic, Sarthak Kamat, Trevor Darrell, and Jitendra Malik. Learning humanoid locomotion over challenging terrain. *arXiv:2410.03654*, 2024.

[Ref18] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. *arXiv preprint arXiv:2106.01345*, 2021.

[Ref19] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016.

[Ref20] Noam Brown and Tuomas Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359:418 – 424, 2018.

[Ref21] Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sandra Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David J. Wu, Hugh Zhang, and Markus Zijlstra. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378:1067 – 1074, 2022.

[Ref22] AI achieves silver-medal standard solving International Mathematical Olympiad problems.

[Ref23] Florian Bordes, Richard Yuanzhe Pang, Anurag Ajay, Alexander C Li, Adrien Bardes, Suzanne Petryk, Oscar Mañas, Zhiqiu Lin, Anas Mahmoud, Bargav Jayaraman, Mark Ibrahim, Melissa Hall, Yunyang Xiong, Jonathan Lebensold, Candace Ross, Srihari Jayakumar, Chuan Guo, Diane Bouchacourt, Haider Al-Tahan, Karthik Padthe, Vasu Sharma, Hu Xu, Xiaoqing Ellen Tan, Megan Richards, Samuel Lavoie, Pietro Astolfi, Reyhane Askari Hemmat, Jun Chen, Kushal Tirumala, Rim Assouel, Mazda Moayeri, Arjang Talattof, Kamalika Chaudhuri, Zechun Liu, Xilun Chen, Quentin Garrido, Karen Ullrich, Aishwarya Agrawal, Kate Saenko, Asli Celikyilmaz, and Vikas Chandra. An introduction to vision-language modeling. *arXiv [cs.LG]*, May 2024.