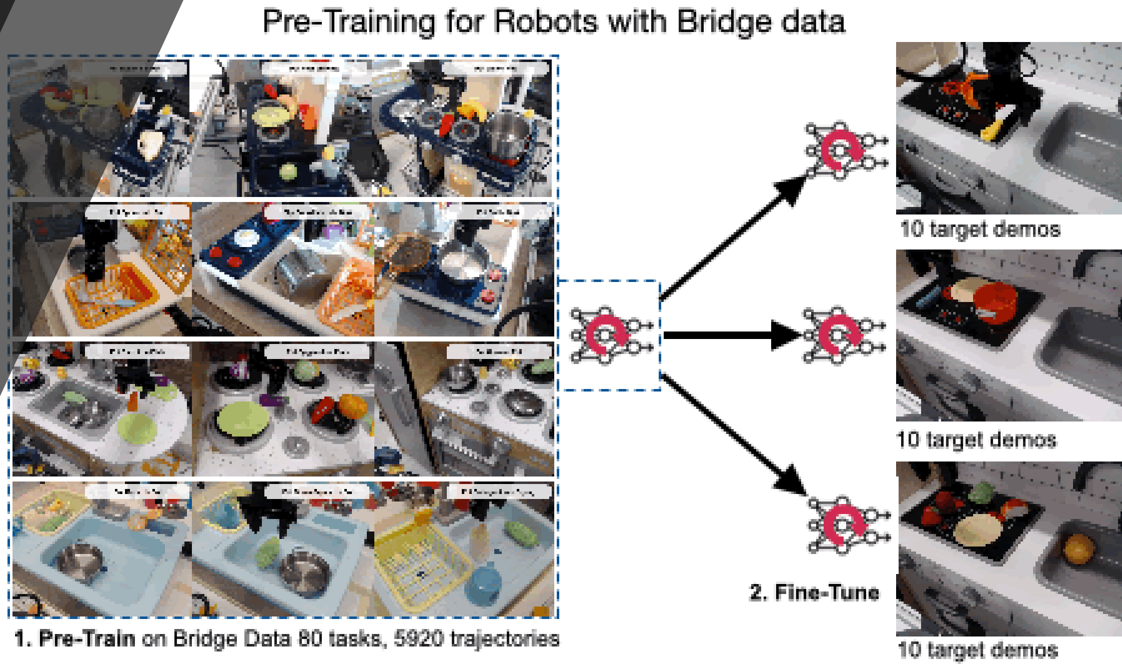
Foundation Models for Reinforcement Learning and Robotics

ROBIN-seminar March 2nd 2023

Kai Olav Ellefsen



1. What are Foundation Models?

The name «Foundation Model»

- Popularized in a 2021 paper from Stanford as they launched CRFM
- Some controversy – not everyone agrees this is a good name
- Other names for the same: Large (Language) Model (LLM), Large Pre-Trained Model

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models *to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.*

[1]Corresponding author: pliang@cs.stanford.edu

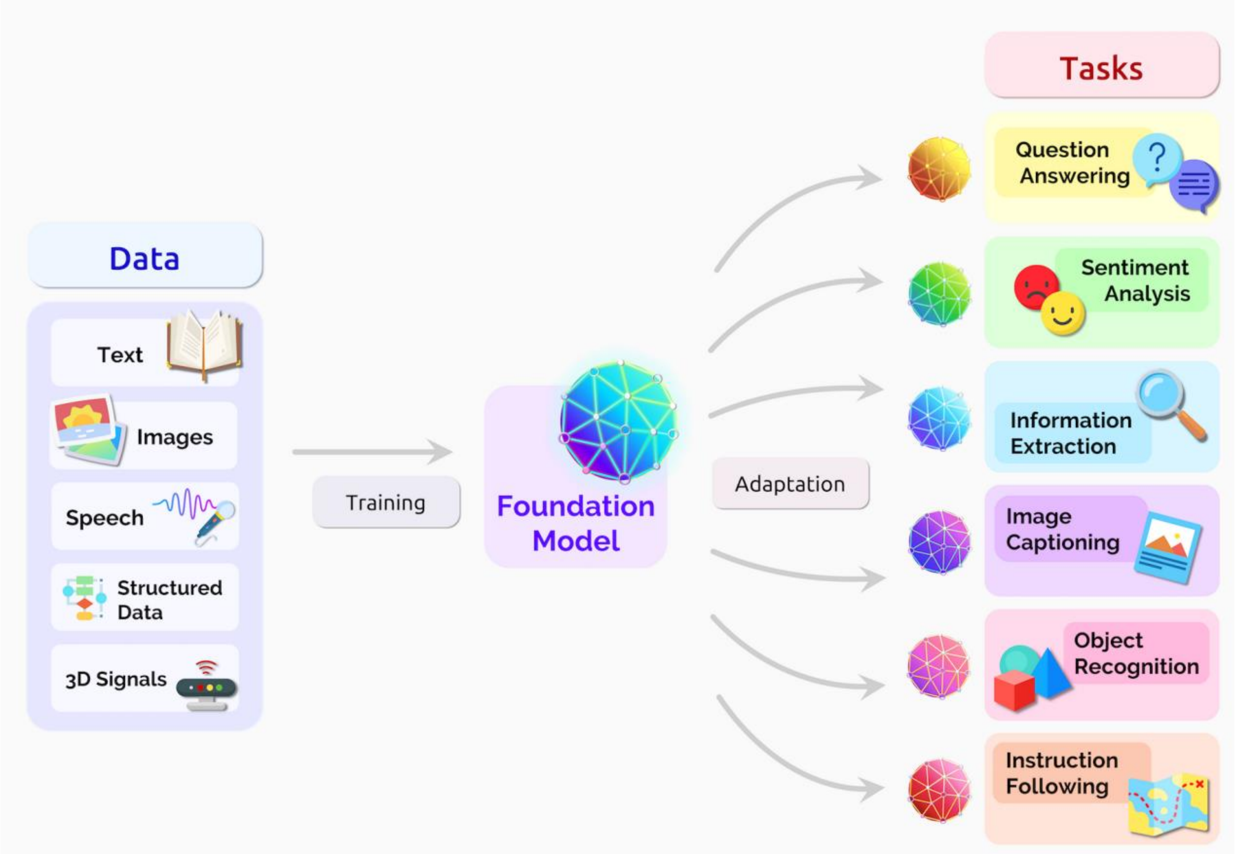
*Equal contribution.

Definition

”A foundation model is any model that is **trained on broad data** (generally using **self-supervision** at scale) that can be **adapted (e.g., fine-tuned)** to a wide range of **downstream tasks**; current examples include BERT [Devlin et al. 2019], GPT-3 [Brown et al. 2020], and CLIP [Radford et al. 2021]”

Common Features of FMs

- Trained using **unsupervised or self-supervised** learning
- Deep Neural Network model with **very many parameters**
- Intended to be a «foundation», that doesn't solve a particular task, but where more **fine-tuning** can give desired specialized abilities



Self-supervised Learning

- Ideally, we would train our FM in a supervised manner
- But we need huge amounts of data. Cannot afford that many labels
- Self-supervised Learning: Get the labels from the data itself.

a red delicious apple

a black office chair

a corgi dog

Self-supervised Learning - Examples

Text: Given a long text sequence as input – predict the most likely next word

Images: Given a part of an image, predict what is in another part

Video: Given 5 seconds of video, predict the next frame

Fine-tuning

Som et lite eksperiment ga jeg en i NORA – The Norwegian Artificial Intelligence Research Consortium tillatelse til å laste meg inn i en slik AI-algoritme. Det heter "fine tuning". Så nå har vi et program som forstår hvem Klas Pettersen er, og som kan gjenskape meg i hvilken som helst situasjon. Jeg vet ikke om det var så lurt ...

Source: Klas Pettersen, Twitter

Why can FMs be Useful for Robotics/ Reinforcement Learning?

Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos

Bowen Baker[*†]
bowen@openai.com

Ilge Akkaya[*†]
ilge@openai.com

Peter Zhokhov[*†]
peterz@openai.com

Joost Huizinga[*†]
joost@openai.com

Jie Tang[*†]
jietang@openai.com

Adrien Ecoffet[*†]
adrien@openai.com

Brandon Houghton[*†]
brandon@openai.com

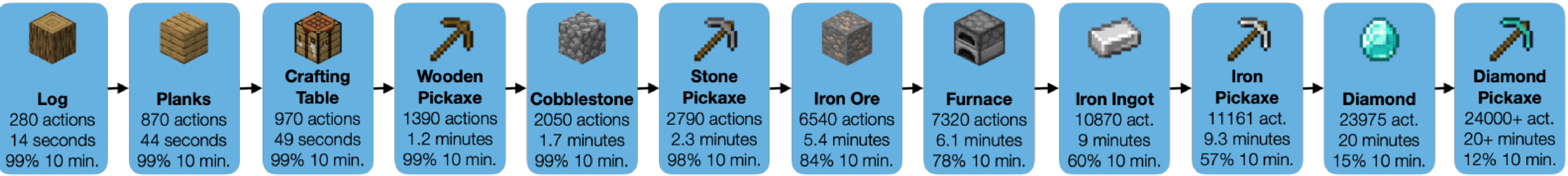
Raul Sampedro[*†]
raulsamg@gmail.com

Jeff Clune[*†‡]
jclune@gmail.com

What the Agents were Trying to Learn

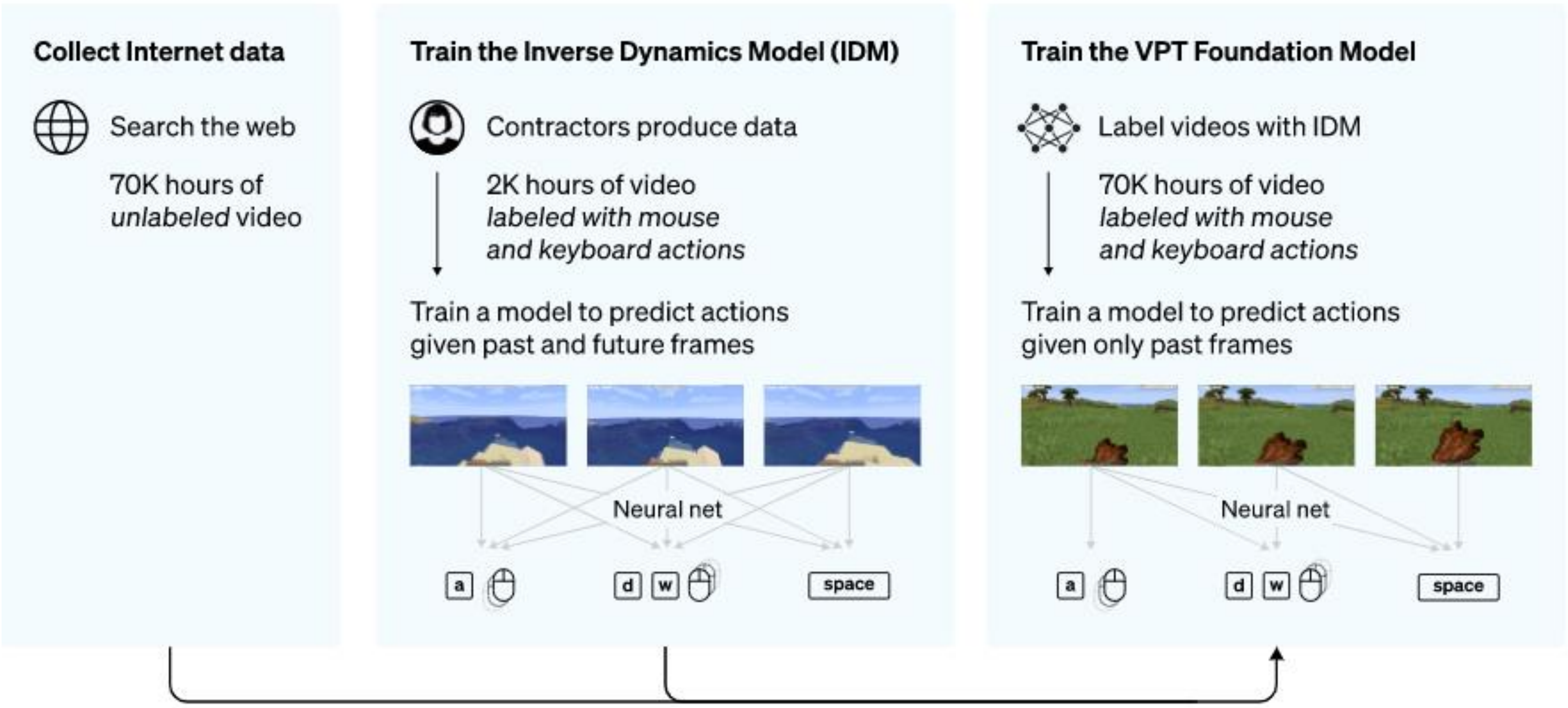
The Point: This is a Very Complex Task

- Requiring very many actions to be taken in the right order
- Learned directly from observing videos of gameplay
- Perhaps the most complex task I've ever seen an RL agent learn?



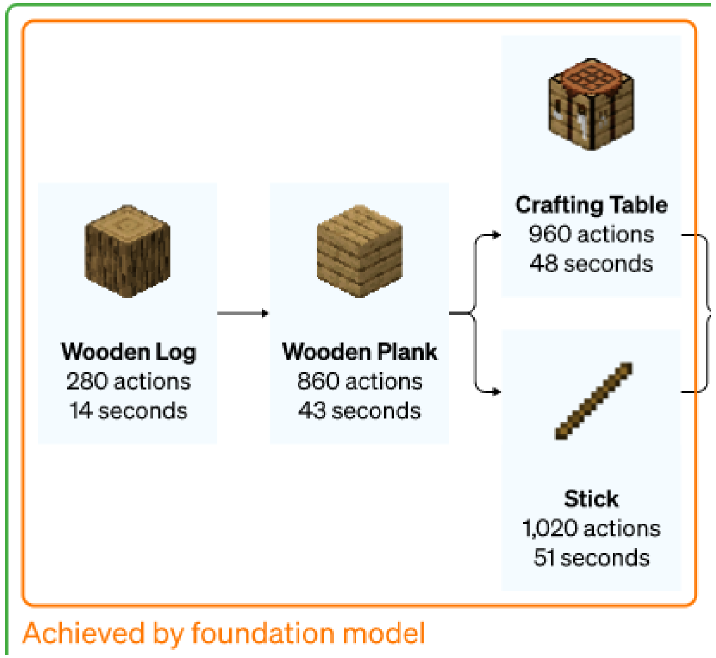
The Idea: Learn a Foundation Model of Basic Minecraft Behaviors

- **YouTube is full of Minecraft videos**. We can learn about how the Minecraft world works from them
- Problem: **This is unlabelled data**. Not too useful since we can’t learn how to act from it.
  - That is, we don’t know what button the player pressed. So we can’t learn relationship between actions (button presses) and outcomes.
- Solution: A form of **semi-supervised learning**, using large amounts of unlabelled and small amount of expensive labelled data.



VPT method overview

So, the Foundation Model is Ready – What can it do?



- This is as expected. The FM is a **foundation** – it can do perfrom **many basic** Minecraft behaviors.
- But as other FM's, it uses fine-tuning to learn more complex behaviors
- Training time: 9 days on 720 V100 GPUs
  - Luckily, you only train an FM once – and fine-tuning is much faster.

The FM Learned Many of the Skills Present in the Training Data

Swimming (zero-shot)

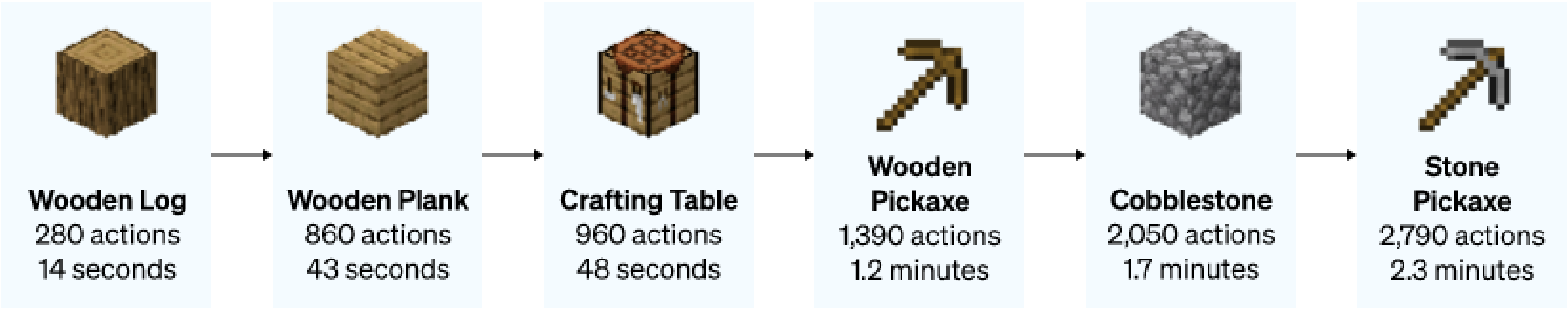

Hunting animals (zero-shot)


Eating food (zero-shot)


Pillar jumping (zero-shot)

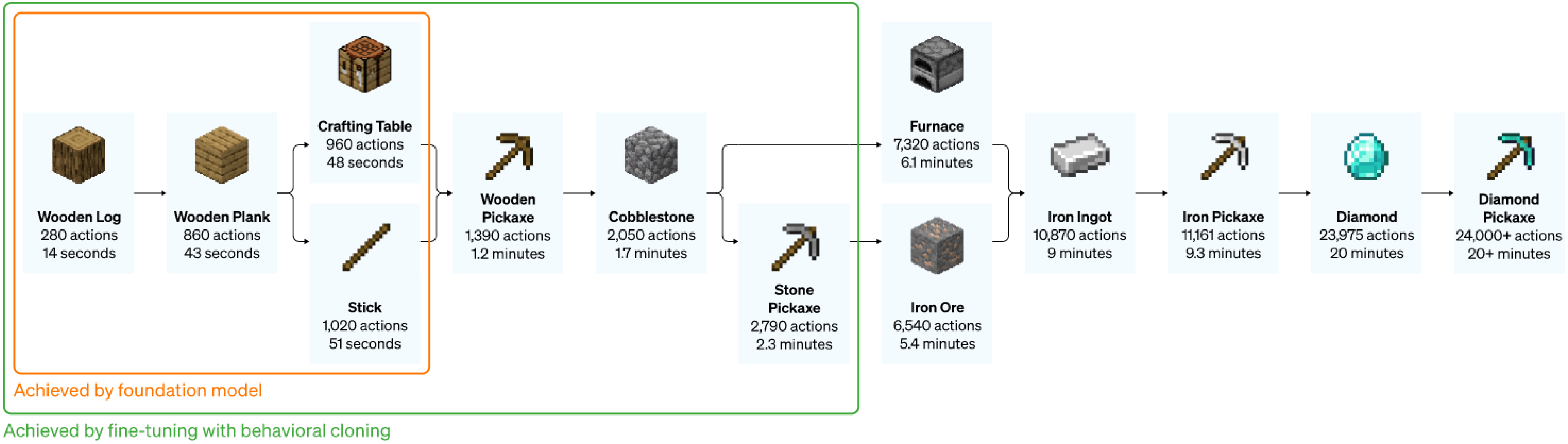
How to Fine-Tune? Option 1: More contractor data

- OpenAI asked human contractors to play some more (10 min per person), but this time with a very specific goal: Build a house
- Fine-tuning the FM with this data allowed it to learn more complex behaviors



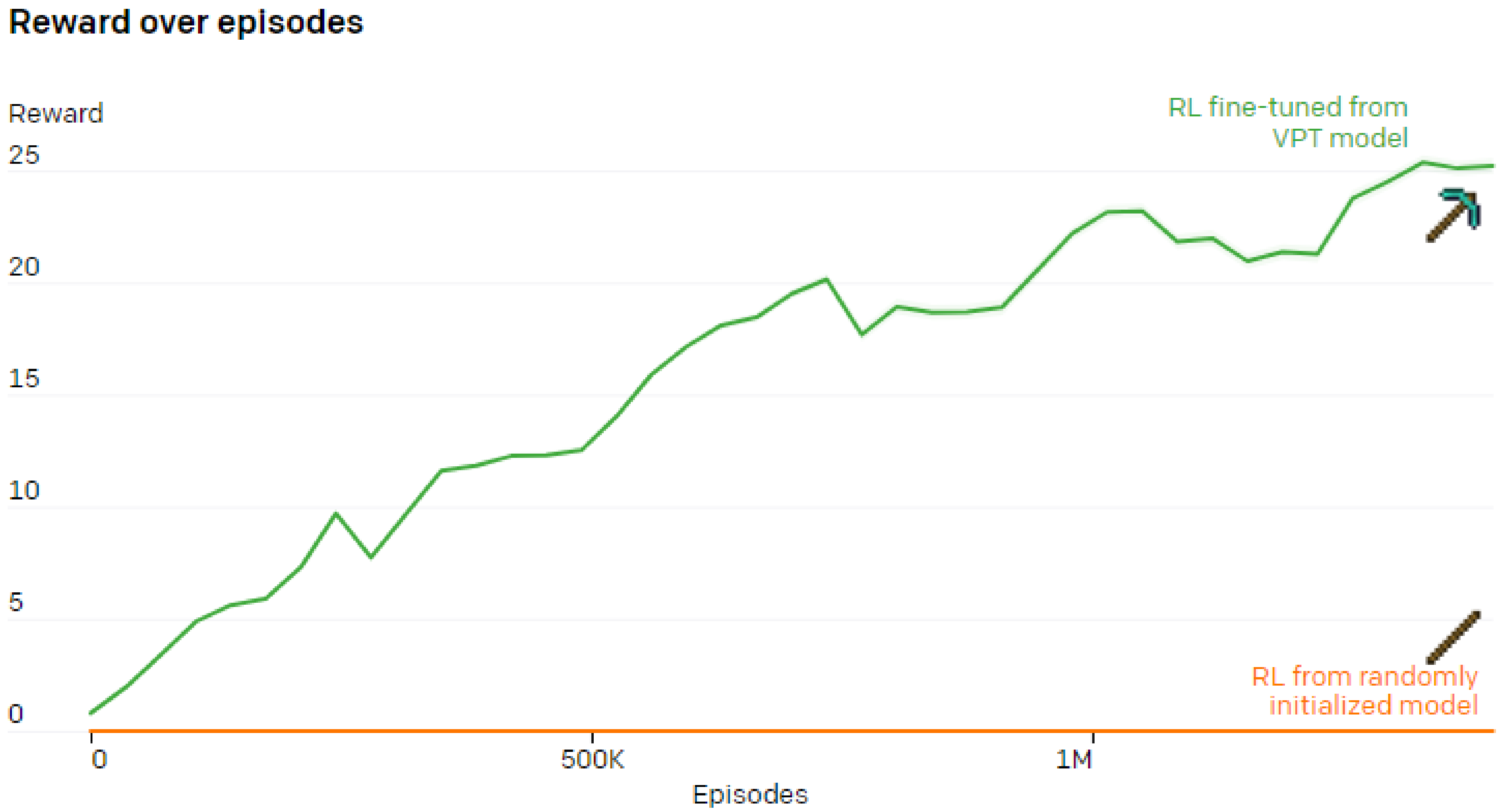
How to Fine-Tune? Option 2: Reinforcement Learning

- The FM is a great starting-point for Reinforcement Learning
- Typically, RL starts with **randomly exploring** the possible actions
  - Minecraft is so complex that randomly exploring will not get you far
- The FM instead «knows» how to behave in Minecraft – so we can **explore only the behaviors that make a certain sense**.
- How do we reach a certain behavior? Give **a reward** for reaching it.
- Here: Reward all steps on the way to the diamond pickaxe

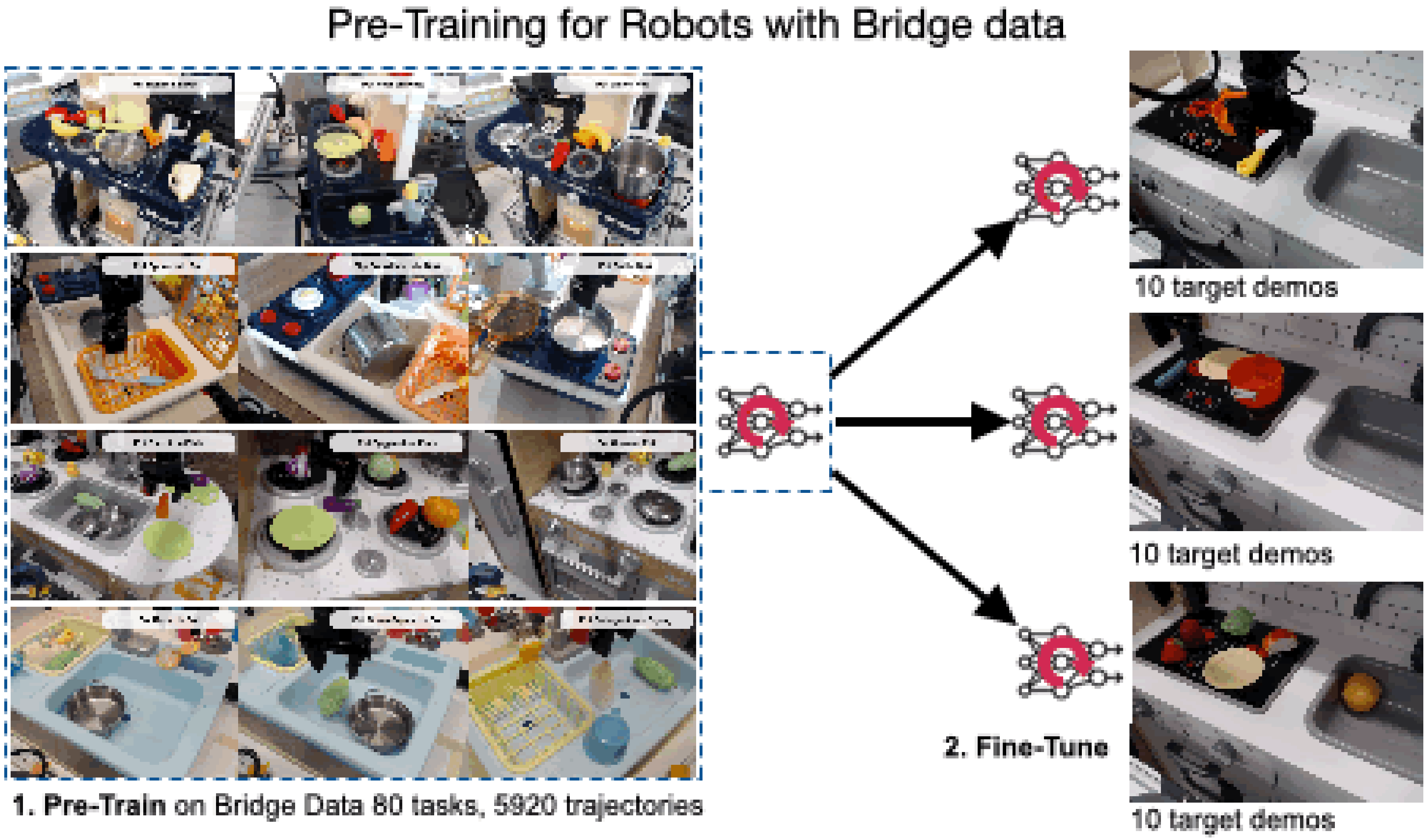


- Result: After training with RL, **2.5% of runs** are able to make the diamond pickaxe
- May sound low, but this was the first time an AI model achieved this
  - For a human it takes 20 minutes and 24.000+ actions!!
- 6 days on 80 GPUs! A lot, but far less than the FM training.

The Importance of the FM



Towards Foundation Models for Robots



https://sites.google.com/view/ptr-robotlearning

Aviral Kumar, Anikait Singh*, Frederik Ebert*, Yanlai Yang, Chelsea Finn, Sergey Levine*

The person who has perhaps thought most about FM’s for robots

Perhaps the person who has thought most about FM’s for robots

Sergey Levine

Associate Professor, UC Berkeley, EECS

Address:
Rm 8056, Berkeley Way West
2121 Berkeley Way
Berkeley, CA 94704

Email:
prospective students: please read this before contacting me.
svlevine AT eecs.berkeley.edu

Follow @svlevine

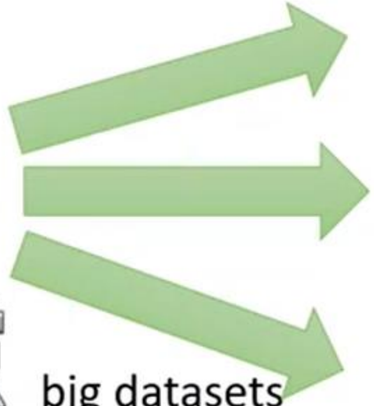
Some things we want in a Robotic Foundation Model

Many robots
Many environments
Many tasks
Goal-conditioned
Fine-tunable

Source: Sergey Levine

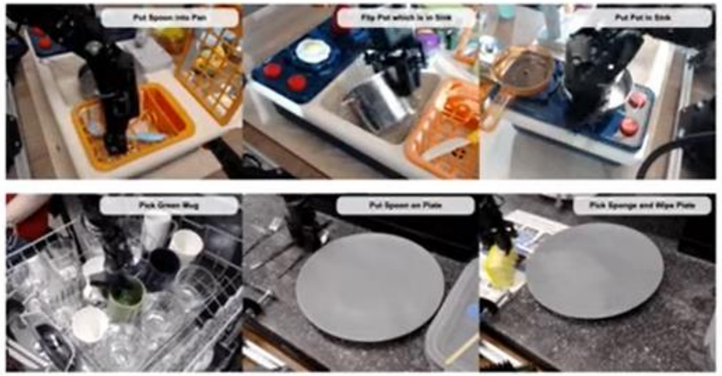
https://www.youtube.com/watch?v=7lDS4j9v1Os&ab_channel=RAIL

Bridge data



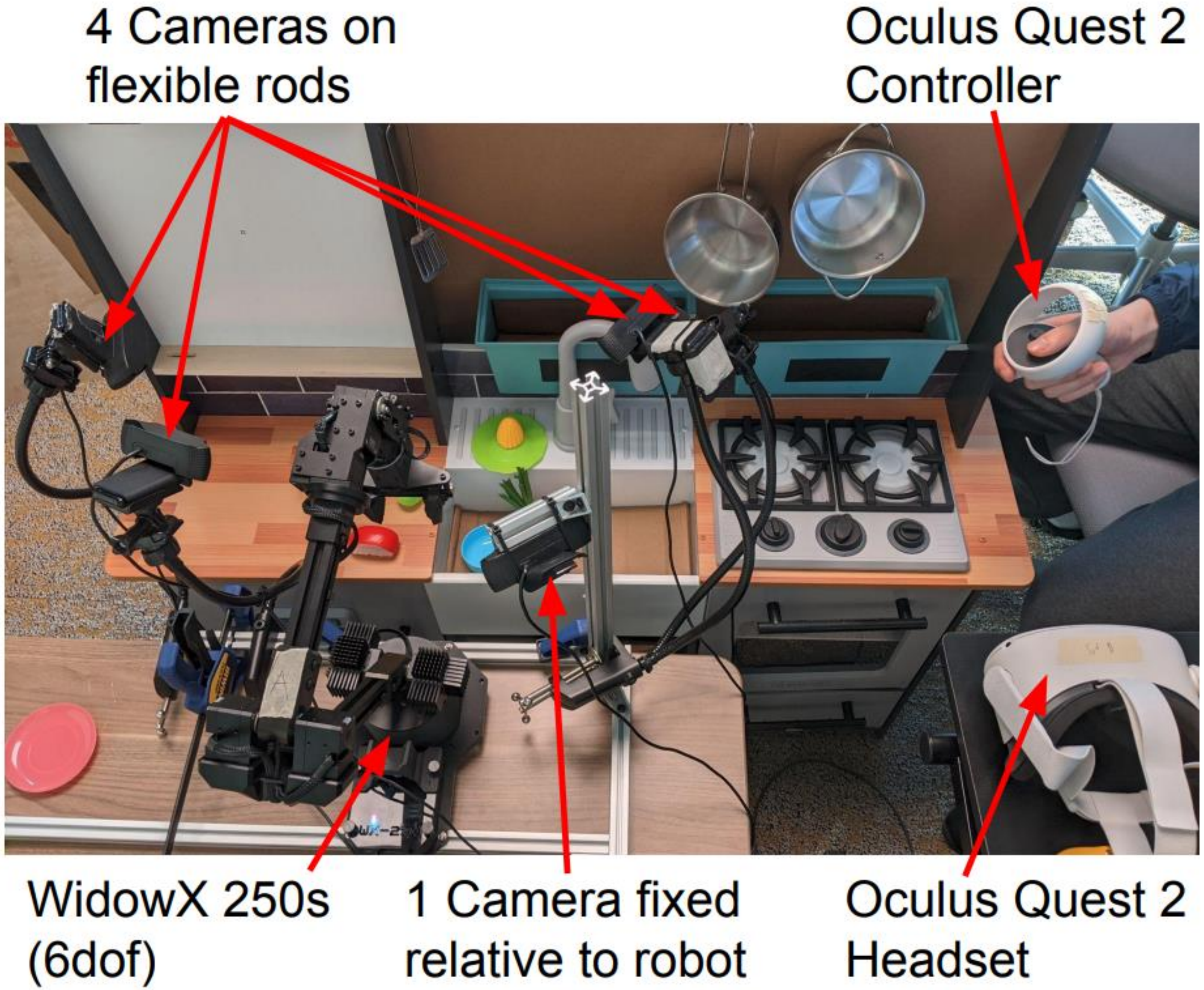
big datasets from past interaction

- 7,000+ demonstrations
- 10 environments
- 70 tasks
- Designed to be reusable by other researchers in **new domains** and for **new tasks**

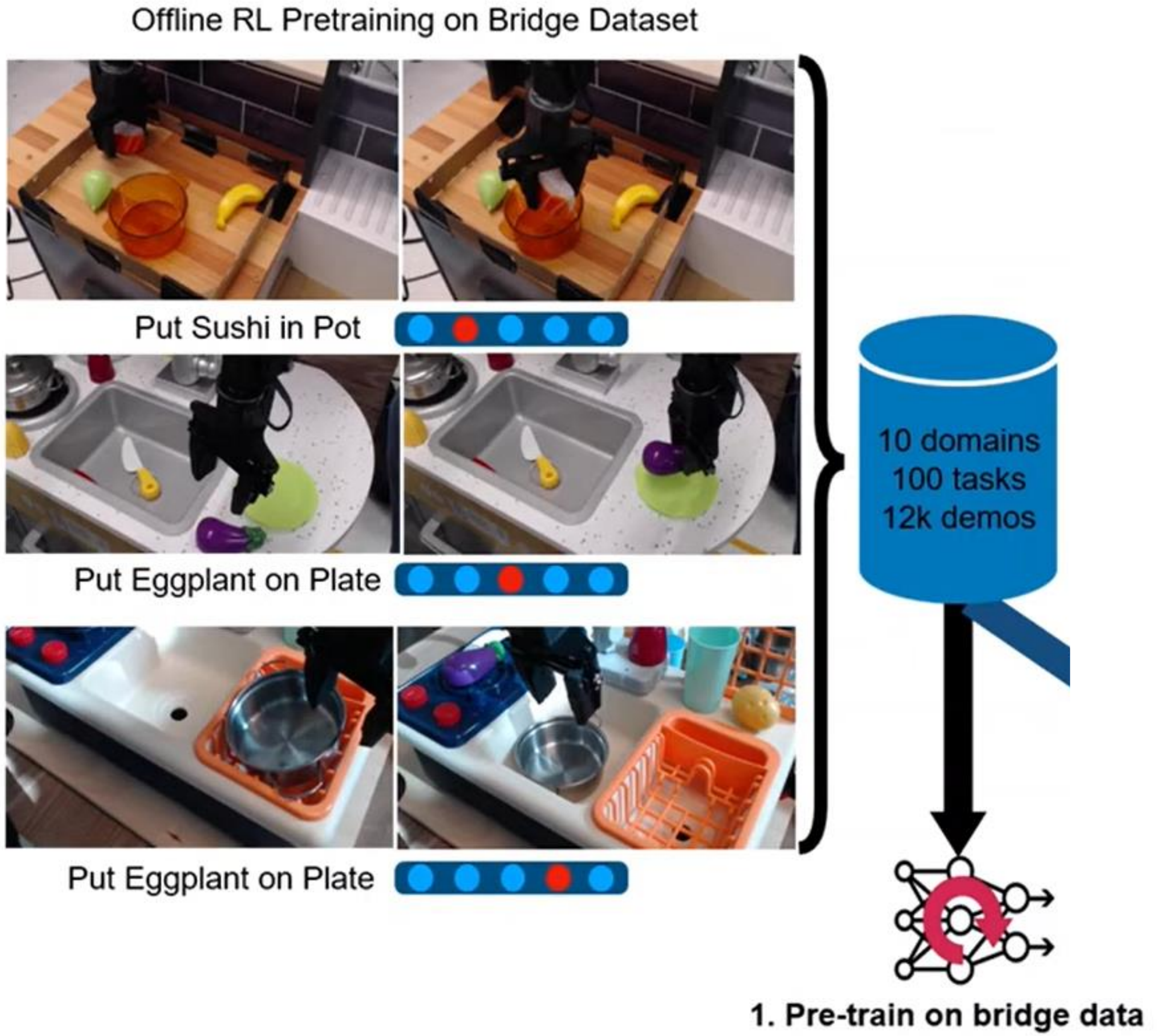


Frederik Ebert, Yanlai Yang, Karl Schmeckpeper, Bernadette Bucher, Georgios Georgakis, Kostas Daniilidis, Chelsea Finn, Sergey Levine.
Bridge Data: Boosting Generalization of Robotic Skills with Cross-Domain Datasets. RSS 2022.

How the data was collected

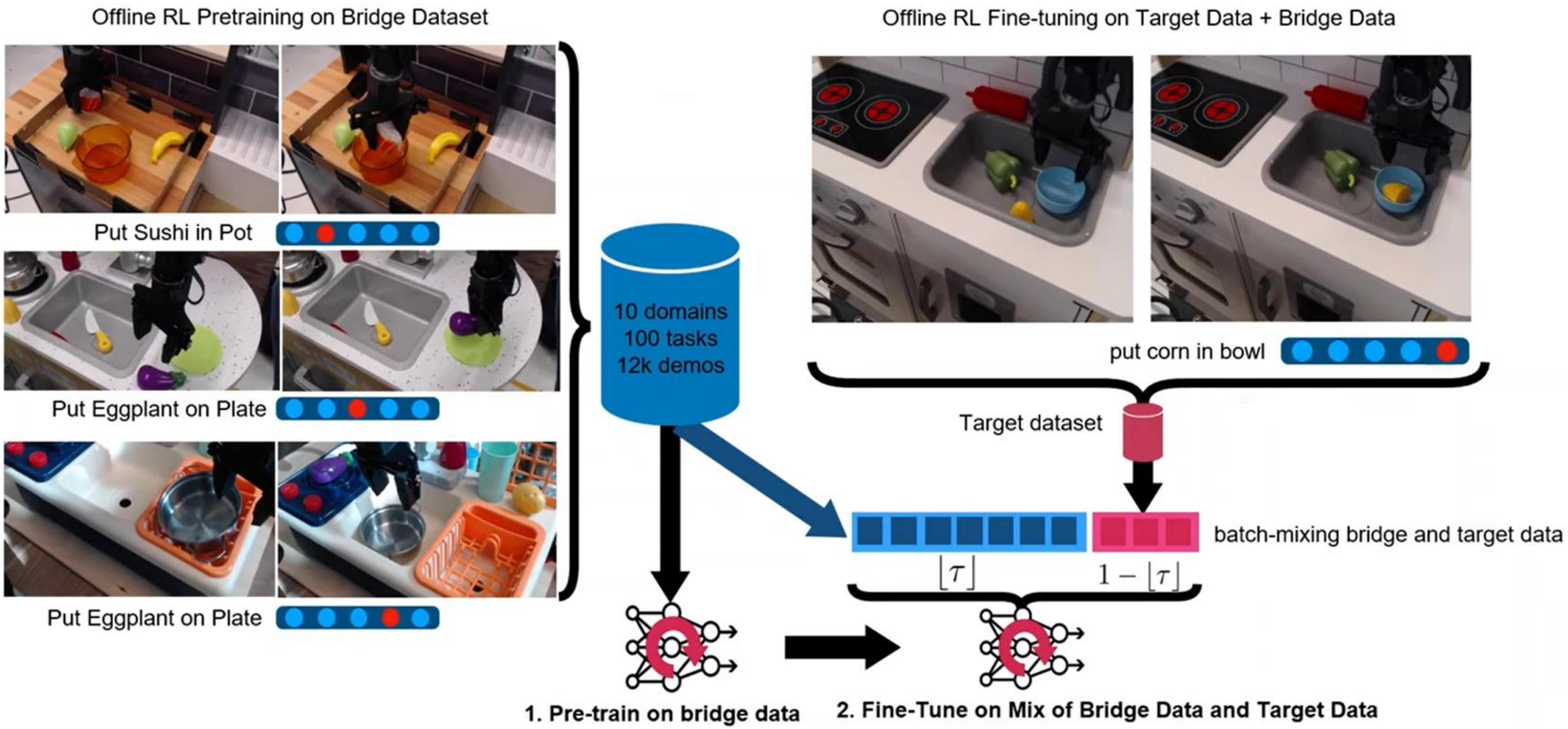


Pretraining on bridge data with offline RL (**PTR**)



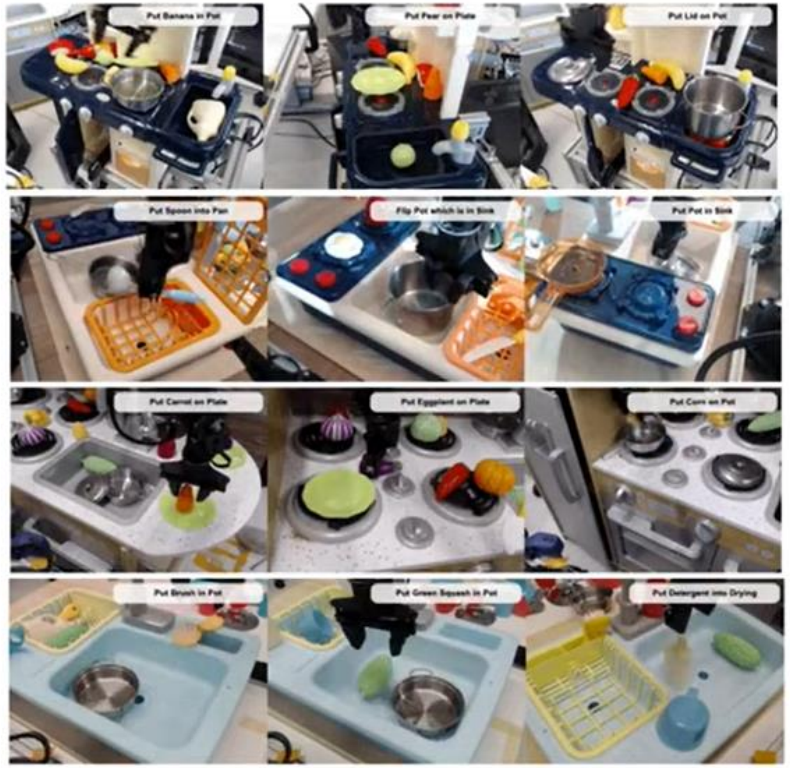
Aviral Kumar, Frederik Ebert, Yanlai Yang, Anikait Singh, Chelsea Finn, Sergey Levine. **Pre-Training for Robots: Offline RL Enables Learning New Tasks from a Handful of Trials.** 2022.

Pretraining on bridge data with offline RL (PTR)



Aviral Kumar, Frederik Ebert, Yanlai Yang, Anikait Singh, Chelsea Finn, Sergey Levine. **Pre-Training for Robots: Offline RL Enables Learning New Tasks from a Handful of Trials.** 2022.

PTR (Pre-Training for Robots) Results

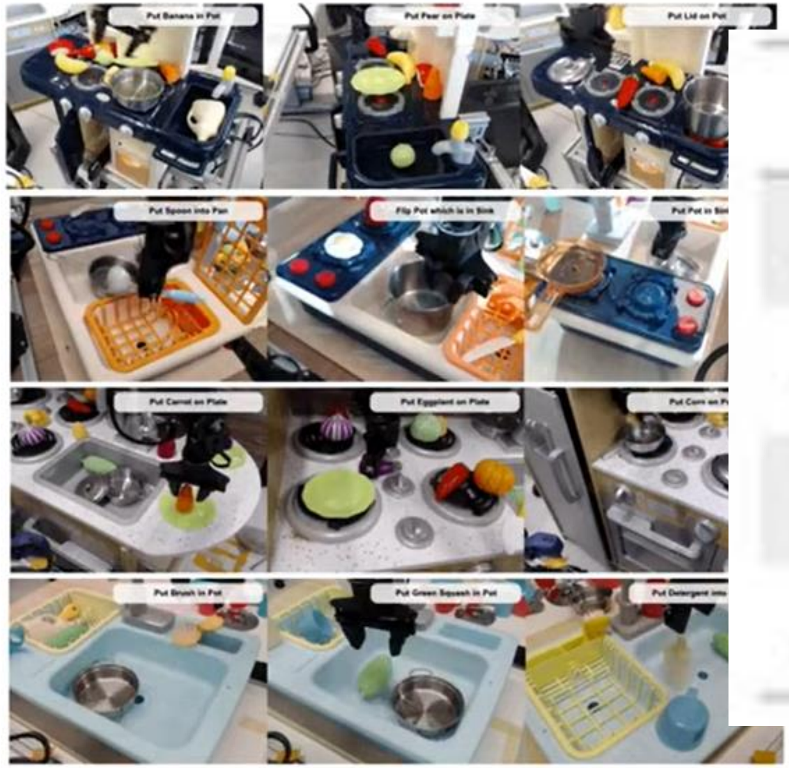


learning **entirely new** skills after pretraining on the bridge dataset

		BC finetuning			Joint training		Target data only		Pre-train. rep. + BC finetune	
Task	**PTR (Ours)**	**BC (fine.)**	**Autoreg. BC**	**BeT**	**COG**	**BC**	**CQL**	**BC**	**R3M**	**MAE**
Take croissant from metal bowl	**7/10**	3/10	5/10	1/10	4/10	4/10	0/10	1/10	1/10	3/10
Put sweet potato on plate	**7/20**	1/20	1/20	0/20	0/20	0/20	0/20	0/20	0/20	1/20
Place knife in pot	**4/10**	2/10	2/10	0/10	1/10	3/10	3/10	0/10	0/10	0/10
Put cucumber in pot	**5/10**	0/10	1/10	0/10	2/10	1/10	0/10	0/10	0/10	0/10

Aviral Kumar, Frederik Ebert, Yanlai Yang, Anikait Singh, Chelsea Finn, Sergey Levine. **Pre-Training for Robots: Offline RL Enables Learning New Tasks from a Handful of Trials.** 2022.

PTR (Pre-Training for Robots) Results



- ✗ Many robots
- ✓ Many environments
- ✓ Many tasks
- ✗ Goal-conditioned
- ✓ Fine-tunable

		BC finetuning			Joint training		Target data only		Pre-train. rep. + BC finetune	
Task	**PTR (Ours)**	**BC (fine.)**	**Autoreg. BC**	**BeT**	**COG**	**BC**	**CQL**	**BC**	**R3M**	**MAE**
Take croissant from metal bowl	**7/10**	3/10	5/10	1/10	4/10	4/10	0/10	1/10	1/10	3/10
Put sweet potato on plate	**7/20**	1/20	1/20	0/20	0/20	0/20	0/20	0/20	0/20	1/20
Place knife in pot	**4/10**	2/10	2/10	0/10	1/10	3/10	3/10	0/10	0/10	0/10
Put cucumber in pot	**5/10**	0/10	1/10	0/10	2/10	1/10	0/10	0/10	0/10	0/10

Aviral Kumar, Frederik Ebert, Yanlai Yang, Anikait Singh, Chelsea Finn, Sergey Levine. **Pre-Training for Robots: Offline RL Enables Learning New Tasks from a Handful of Trials.** 2022.

Conclusion

- Foundation Models for robotics are beginning to show some potential
- Still lots of work to be done, especially to generalize across robots
- I think FMs are promising, as they allow also labs without huge amounts of GPUs or lab assistants to work with SOA models:
  - Big labs build the FM or collect the data, puts it on online
  - Others can use it, fine-tune it to their specific tasks.