

# **DEEPMIND'S GENERAL AI**

A GENERALIST AGENT

a.k.a GATO

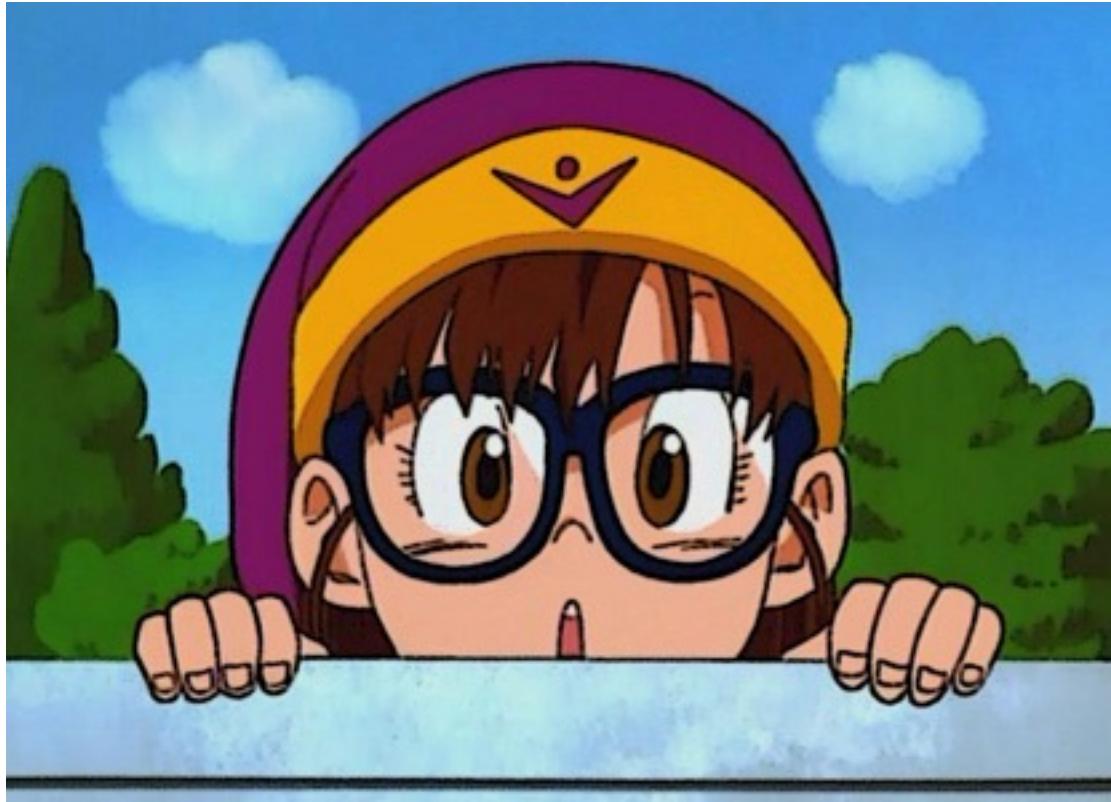


◦ 이 논문을 리뷰한 이유?



Represented by DK Yoon

A Generalist Agent (Deepmind, 2022)



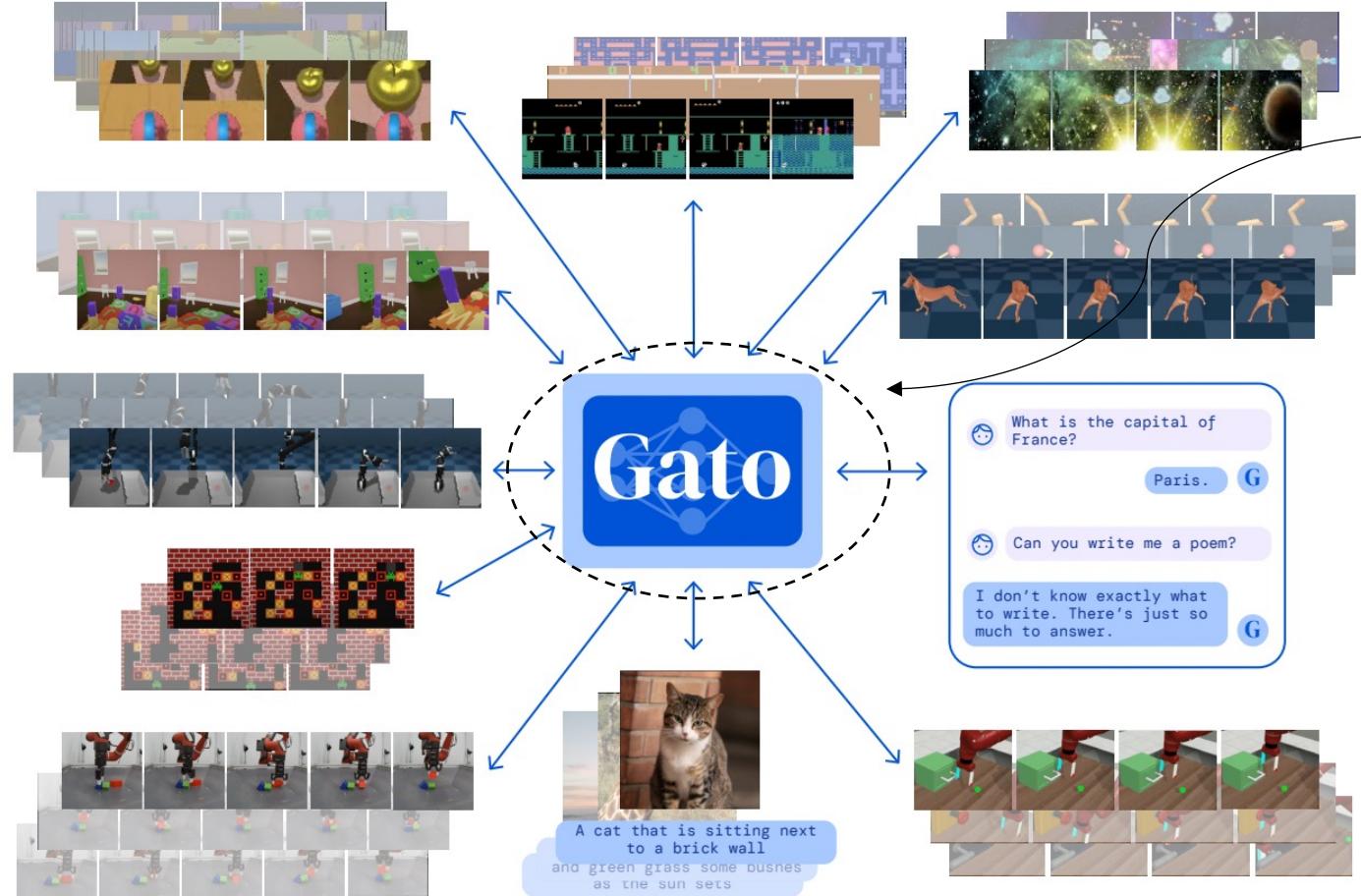
시각/인식

언어

동작

애기(AGI) 가토는 “      “도 가능해요

# Introduction



GATO는 다양한 Task를 다양한 Input 형태, 다양한 Action 형태를 갖는 다양한 Task를 수행할 수 있습니다.

그렇게 만들어진 하나의 모델로 Atari 게임, 이미지에 주석을 달는 것, 채팅, 실제 로봇으로 블록을 쌓는 것 등등이 가능합니다.

Input 형태에 관계없이 토크나이즈 가능하고 이를 통해 의사결정이 있다면 어떤 task든 학습할 수 있습니다!

# Training

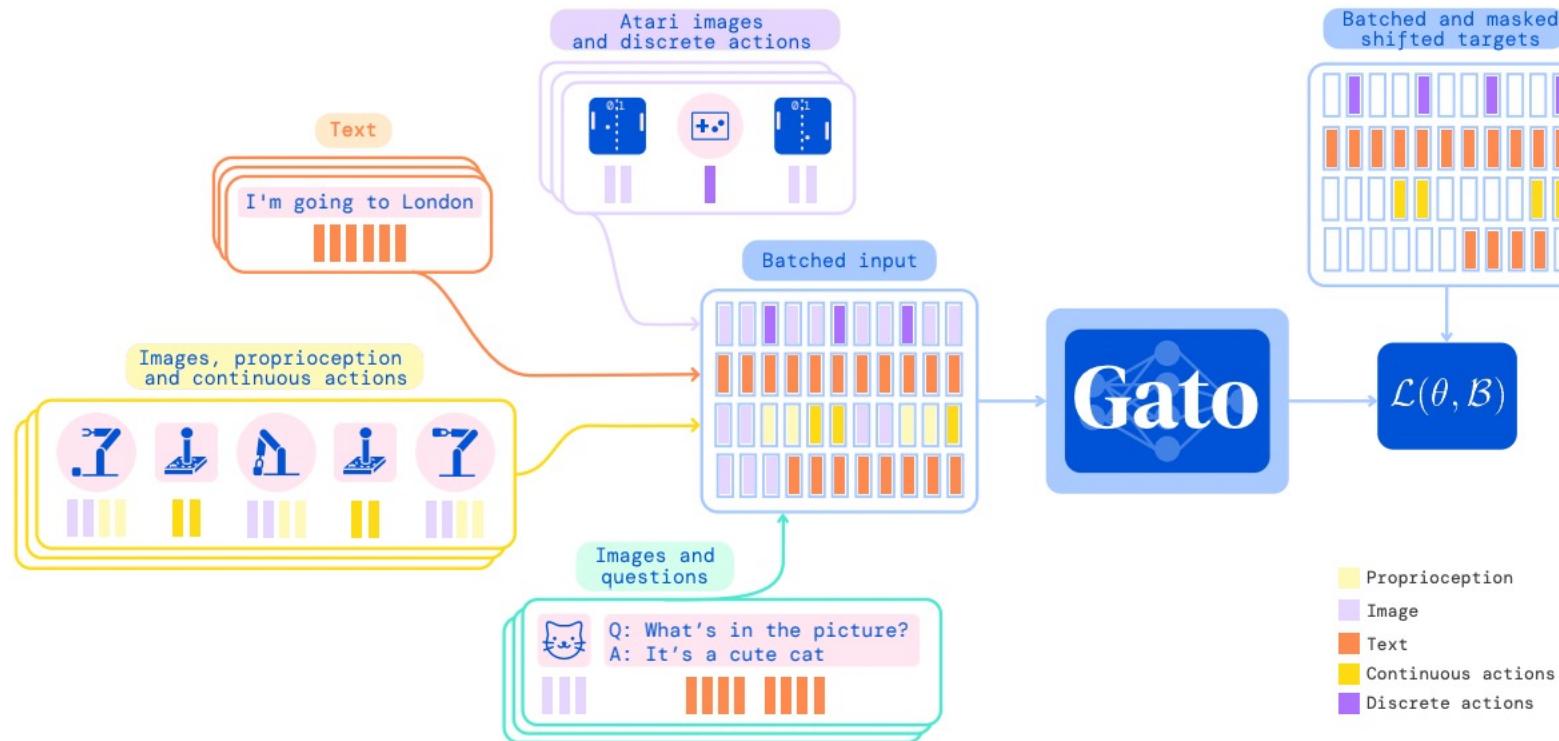
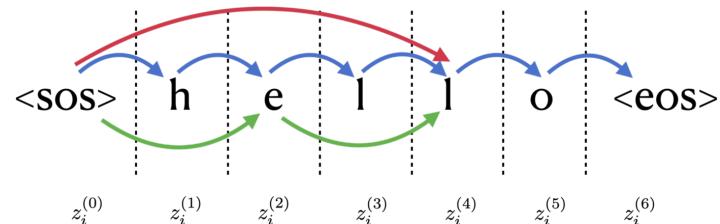


Figure. Gato의 훈련 단계에서 다양한 작업 및 형태의 데이터는

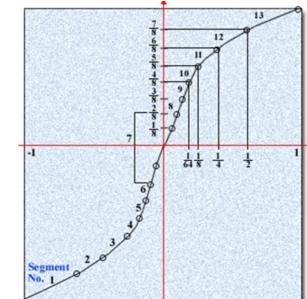
- (1) 시퀀스 모양의 토큰으로 만들어진다.
- (2) 대규모 언어 모델과 유사한 트랜스포머 신경망에 의해 처리됨
- (3) Masking된 타겟과 비교를 통해 손실(Loss)가 계산된다.

# Tokenization

- Text: Sentence Piece [Kudo and Richardson, 2018] 방법을 이용하여  $[0, 32000]$ 로 인코딩
- Image:  $18 \times 18$  패치로 분할 [Dosovitskiy et al., 2020]
- Discrete : 정수  $\in [0, 1024)$ 로 구성된 시퀀스로 변환
- Continuous: Mu-law 인코딩  $\in [-1, 1)$ , Discretize, shifted  $\in [32000, 33024)$



$\mu$ -law Companding:



$$f(x_t) = sign(x_t) \frac{\ln(1 + \mu |x_t|)}{\ln(1 + \mu)}$$

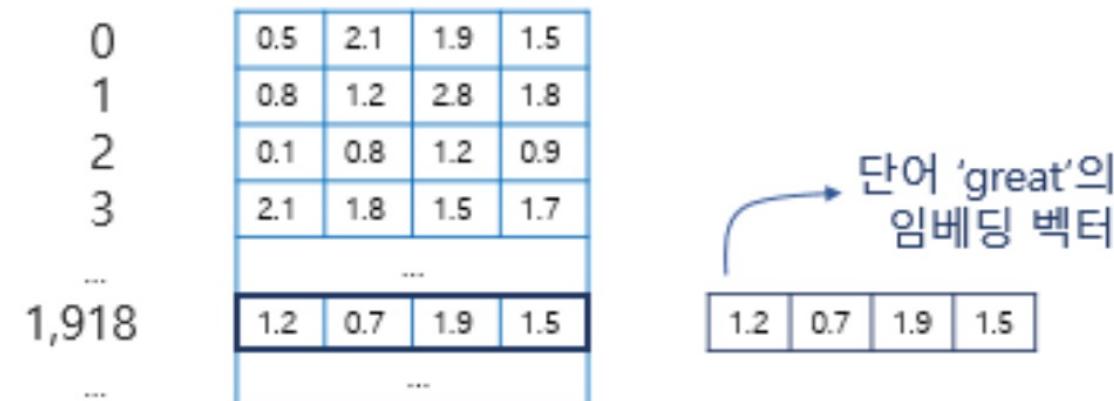
# Embedding (not Image)

---

Tokenization 후, 임베딩 함수가 적용됩니다.  $f(\cdot; \theta_e)$

- Lookup table : 학습된 임베딩 벡터에 대한 룩업 테이블.

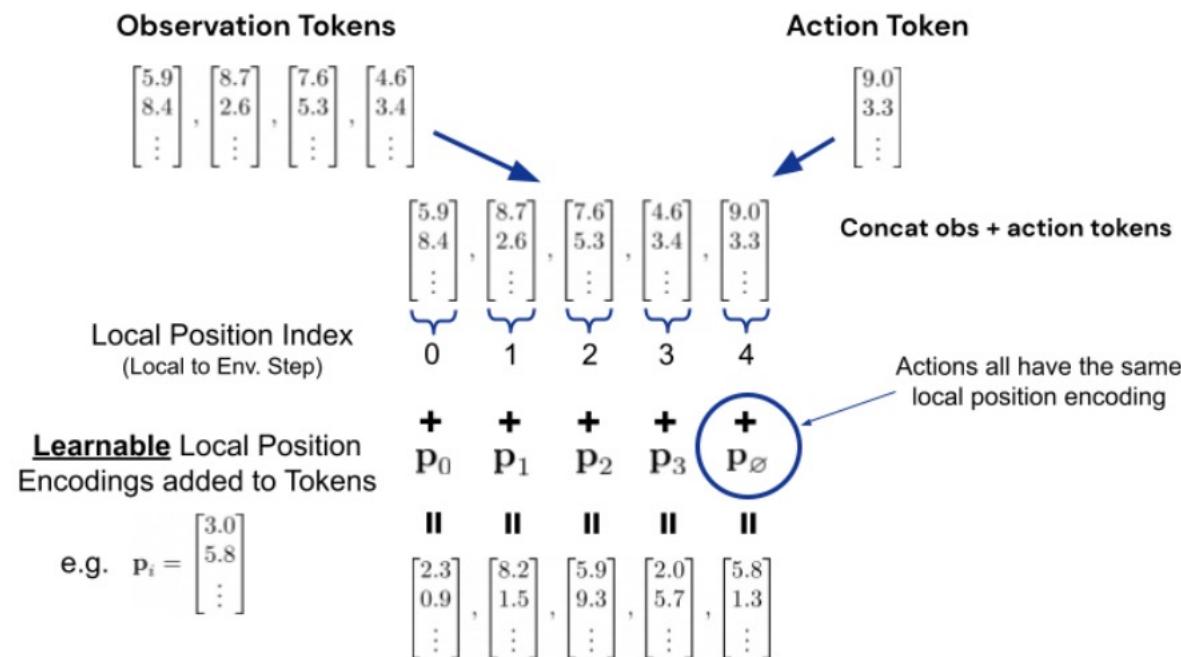
**Integer → lookup Table → Embedding vector**



훈련 과정에서 학습된다.

# Local Position Encoding

- Local Position Index에 대해 Local position encoding이 추가됩니다.
  - (Action은 같은 인코딩 값으로 주어집니다.)



# Embedding (Image)

Tokenization 후, 임베딩 함수가 적용됩니다.  $f(\cdot; \theta_e)$

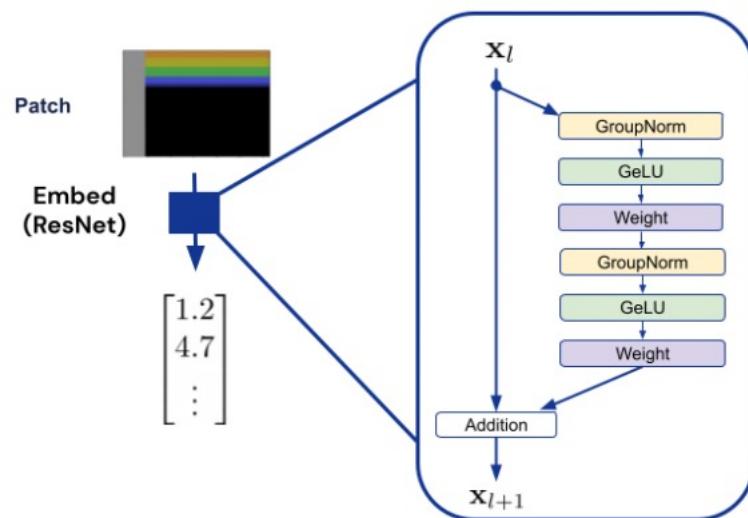


Fig. Resnet Embedding

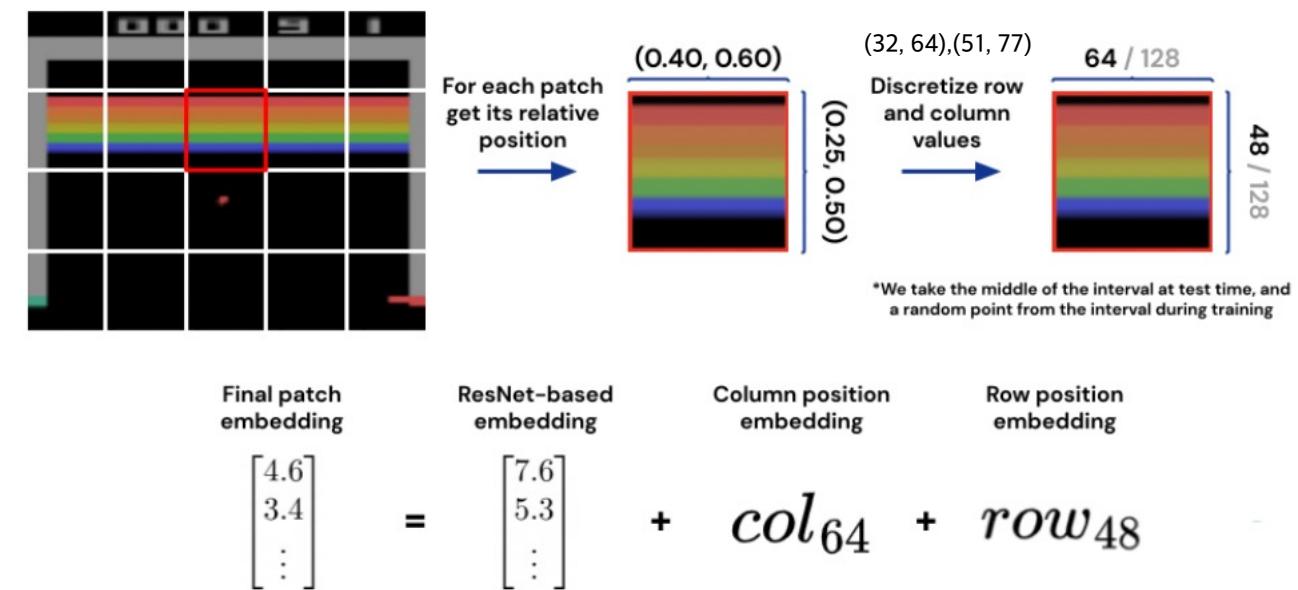


Fig. Patch Position Encoding (patch size 16)

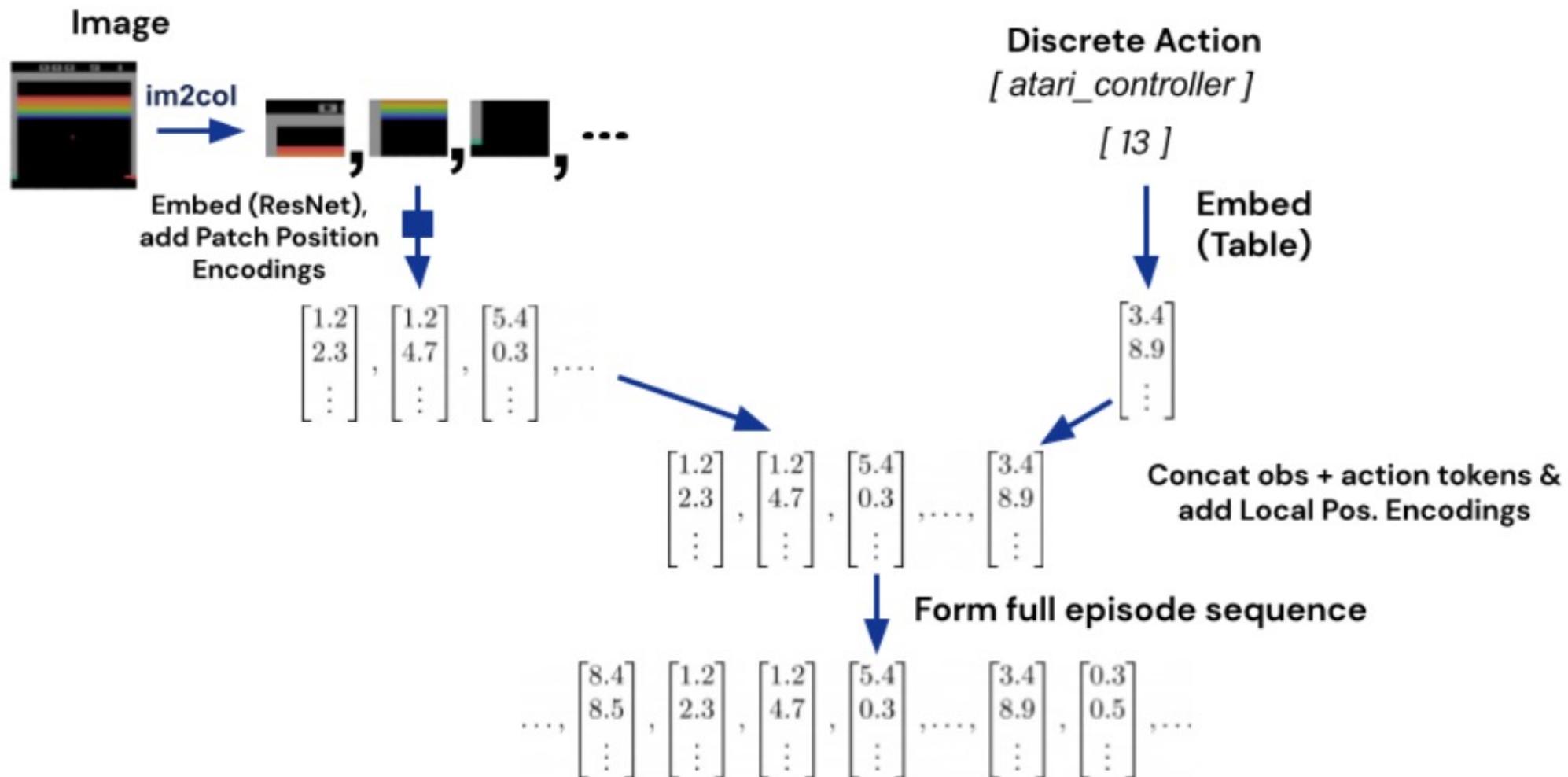
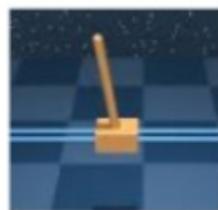


Figure 12 | A visualization of tokenizing and sequencing images and discrete values.



### Proprioception

[*cart\_x, velocity\_x, pole\_x, pole\_y*]

[ 0.3, 1.5, 0.25, -0.4 ]

### Continuous Action

[*cart\_move\_x*]

[ -0.7 ]

↓  
Discretize & Embed (Table)

↓  
Mu-law Encode,  
Discretize & Embed  
(Table)

$$\begin{bmatrix} 5.9 \\ 8.4 \\ \vdots \end{bmatrix}, \begin{bmatrix} 8.7 \\ 2.6 \\ \vdots \end{bmatrix}, \begin{bmatrix} 7.6 \\ 5.3 \\ \vdots \end{bmatrix}, \begin{bmatrix} 4.6 \\ 3.4 \\ \vdots \end{bmatrix}$$

Token  
Embedding

$$\begin{bmatrix} 5.9 \\ 8.4 \\ \vdots \end{bmatrix}, \begin{bmatrix} 8.7 \\ 2.6 \\ \vdots \end{bmatrix}, \begin{bmatrix} 7.6 \\ 5.3 \\ \vdots \end{bmatrix}, \begin{bmatrix} 4.6 \\ 3.4 \\ \vdots \end{bmatrix}, \begin{bmatrix} 9.0 \\ 3.3 \\ \vdots \end{bmatrix}$$

$$\begin{bmatrix} 9.0 \\ 3.3 \\ \vdots \end{bmatrix}$$

Concat obs + action tokens &  
add Local Pos. Encodings

↓  
Form full episode sequence

$$\dots, \begin{bmatrix} 2.3 \\ 8.3 \\ \vdots \end{bmatrix}, \begin{bmatrix} 5.9 \\ 8.4 \\ \vdots \end{bmatrix}, \begin{bmatrix} 8.7 \\ 2.6 \\ \vdots \end{bmatrix}, \begin{bmatrix} 7.6 \\ 5.3 \\ \vdots \end{bmatrix}, \begin{bmatrix} 4.6 \\ 3.4 \\ \vdots \end{bmatrix}, \begin{bmatrix} 9.0 \\ 3.3 \\ \vdots \end{bmatrix}, \begin{bmatrix} 2.3 \\ 0.9 \\ \vdots \end{bmatrix}, \dots$$

Figure 13 | A visualization of tokenizing and sequencing continuous values, e.g. proprioception.

# Training

---

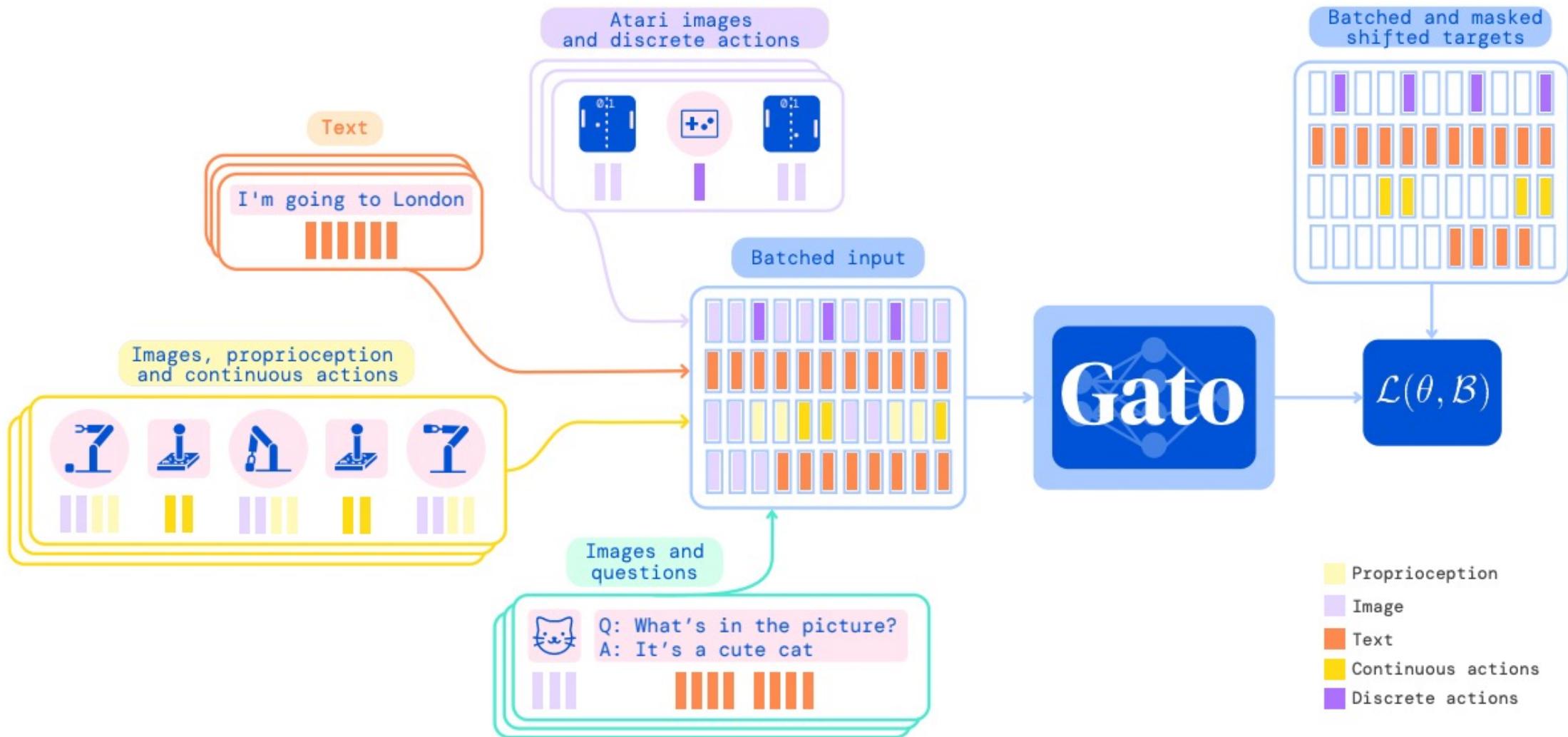
$$p(a, b, c) = p(a|b, c)p(b|c), p(c)$$

- $s_{1:L}$ 과 파라미터  $\theta$ 로 자동회귀모형(Autoregressively model) 방식으로 학습합니다.

$$\log p_\theta(s_1, \dots, s_L) = \sum_{i=1}^L \log p_\theta(s_i | s_1, \dots, s_{i-1})$$

$$\mathcal{L}(\theta, \mathcal{B}) = - \sum_{b=1}^{|\mathcal{B}|} \sum_{l=1}^L m(b, l) \log p_\theta(s_l^{(b)} | s_1^{(b)}, \dots, s_{l-1}^{(b)})$$

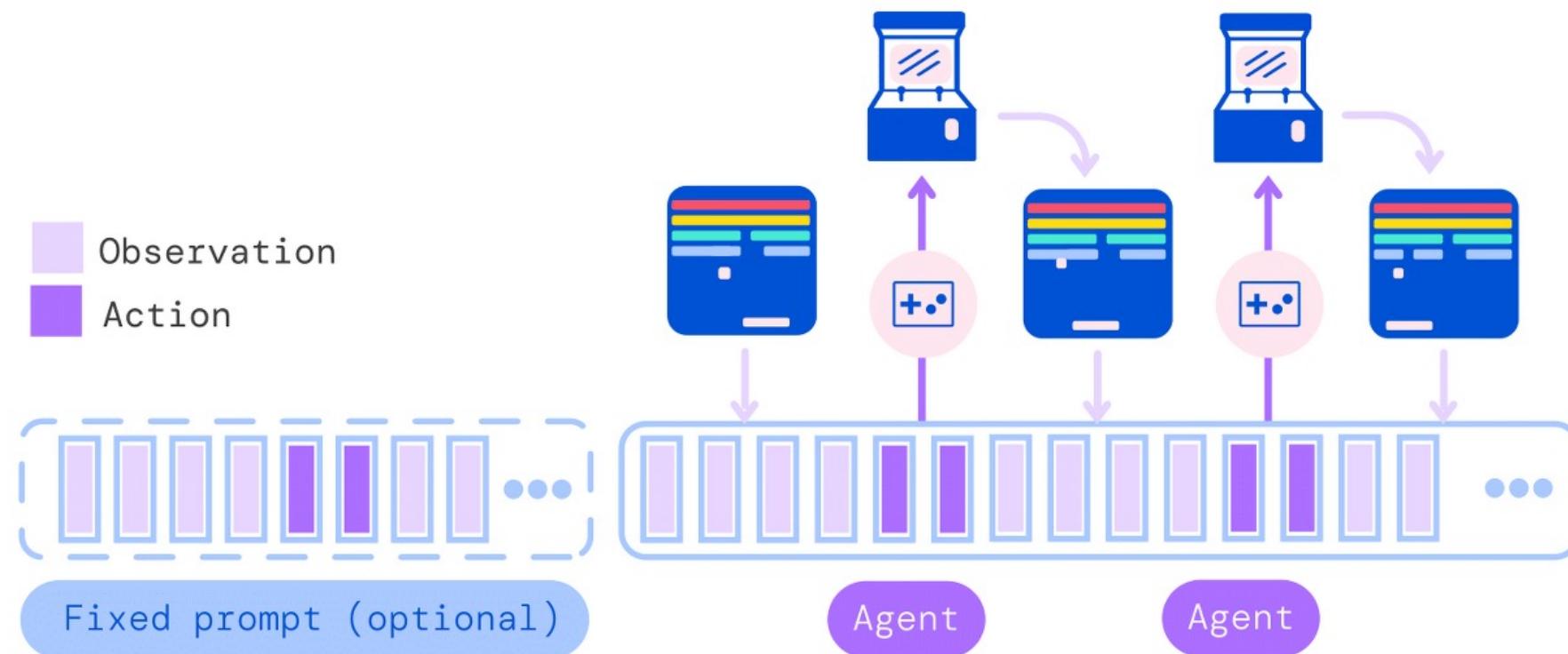
- 네트워크 아키텍처는 다음으로 구성되어 있다.
  - 매개화된 임베딩 함수
  - 시퀀스 모델 (1.2B, decoder만 있는 transformer(24 layers))
- 각 배치는 여러 태스크(환경)의 데이터가 섞여서 들어간다.



# Deployment

---

- Running Gato as a control policy



# Dataset

Total 604!

---

Control environment	Tasks	Episodes	Approx. Tokens	Sample Weight	Vision / language dataset	Sample Weight
DM Lab	254	16.4M	194B	9.35%	MassiveText	6.7%
ALE Atari	51	63.4K	1.26B	9.5%	M3W	4%
ALE Atari Extended	28	28.4K	565M	10.0%	ALIGN	0.67%
Sokoban	1	27.2K	298M	1.33%	MS-COCO Captions	0.67%
BabyAI	46	4.61M	22.8B	9.06%	Conceptual Captions	0.67%
DM Control Suite	30	395K	22.5B	4.62%	LTIP	0.67%
DM Control Suite Pixels	28	485K	35.5B	7.07%	OKVQA	0.67%
DM Control Suite Random Small	26	10.6M	313B	3.04%	VQAV2	0.67%
DM Control Suite Random Large	26	26.1M	791B	3.04%		
Meta-World	45	94.6K	3.39B	8.96%	Total	14.7%
Procgen Benchmark	16	1.6M	4.46B	5.34%		
RGB Stacking simulator	1	387K	24.4B	1.33%		
RGB Stacking real robot	1	15.7K	980M	1.33%		
Modular RL (Mujoco)	38	843K	69.6B	8.23%		
DM Manipulation Playground	4	286K	6.58B	1.68%		
Playroom	1	829K	118B	1.33%		
Total	596	63M	1.5T	85.3%		

\* Sample weight means the proportion of each dataset, on average, in the training sequence batches.

시뮬레이션, 실제, 이미지, Text 등 포함

# Dataset

---

- Control task에 대해, Return이 전문가의 Return의 80%를 넘는 에피소드만 필터링해서 수집합니다.

$$\text{expert return} = \max_{j \in [0, 1, \dots, N-W]} \left( \sum_{i=j}^{j+W-1} \frac{R_i}{W} \right)$$

- $N$ : 해당 task에 대해 수집된 에피소드의 총 개수
- $W$ : window size (we set  $w = \min(1000, 0.1 \times N)$ )
- $R_i$  : 에피소드  $i$ 의 총 Return
- 데이터 수집에 활용된 Agent : Muesli, D4PG, MPO, IMPALA, R2D2, ...

# Experiment Detail

---

## Pretraining

- 1M steps
- AdamW
- Linear Warm up from 1e-7 (15000 steps)
- Cosine decay (1M steps)
- Batch size : 512, Sequence length : 1024
- Stochastic depth

## Fine Tuning

- 5000 steps만 학습
- Adam
- Batch size: 64, Sequence length : 1024
- 100 step마다 평가 (moving average로 성능 평가)

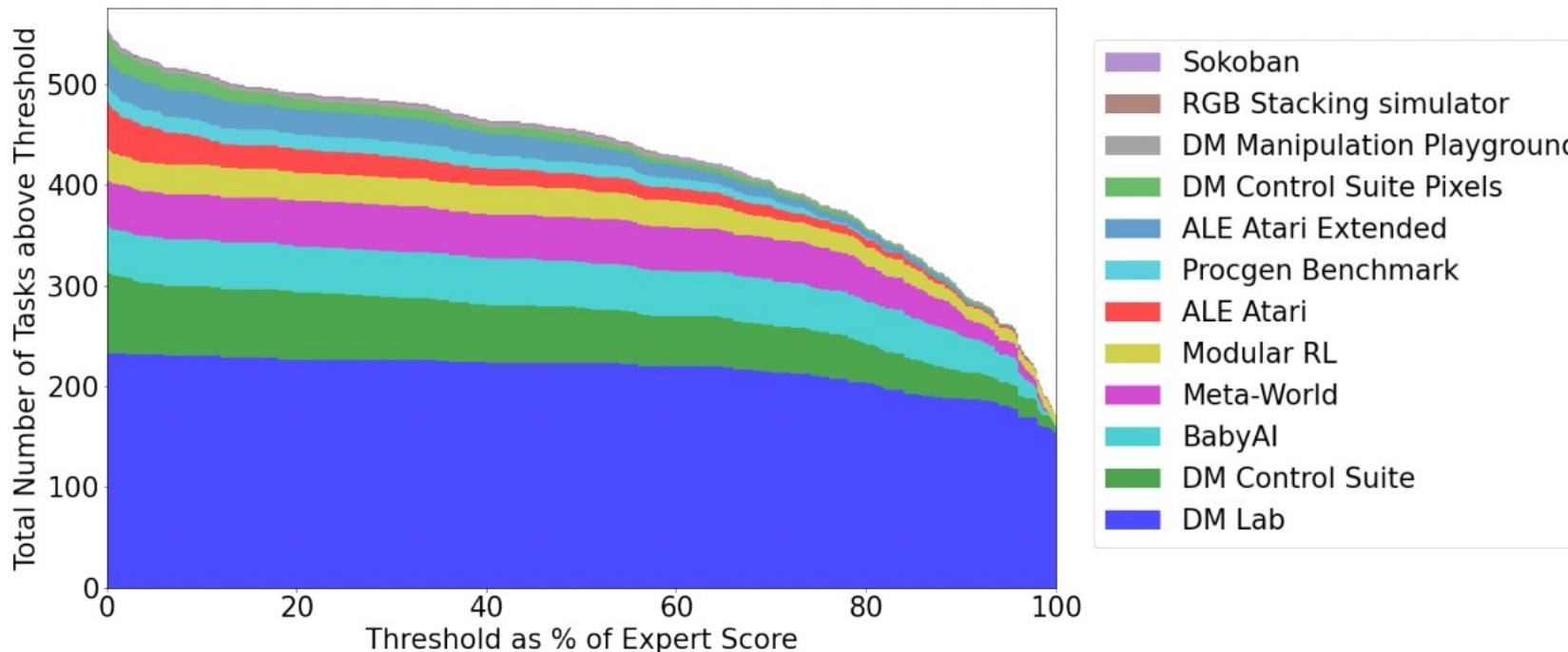
Table 6 | Learning rate schedule hyperparameters for the different model scales.

HYPERPARAMETER	GATO 1.18B	364M	79M
MAXIMUM LEARNING RATE	1E-4	2E-4	1E-4
MINIMUM LEARNING RATE	1E-5	2E-5	1E-5

5e-4..

# Experiment Result

---



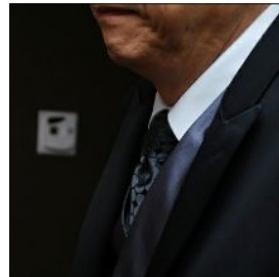
- 604개의 task 중 450개의 성능이 Expert Score 50% 이상이 나옴
- ALE Atari에서 24게임은 사람보다 좋은 성능이 나옴 (RL 44개)

# Experiment Result (Captioning)

---



The colorful ceramic toys are on the living room floor.  
a living room with three different color deposits on the floor  
a room with a long red rug a tv and some pictures



Man standing in the street wearing a suit and tie.  
A man in a blue suit with a white bow tie and black shoes.  
A man with a hat in his hand looking at the camera



A bearded man is holding a plate of food.  
Man holding up a banana to take a picture of it.  
a man smiles while holding up a slice of cake



a group of people that is next to a big horse  
A tan horse holding a piece of cloth lying on the ground.  
Two horses are laying on their side on the dirt.



Man biting a kite while standing on a construction site  
a big truck in the middle of a road  
A truck with a kite painted on the back is parked by rocks.



a white horse with a blue and silver bridle  
A white horse with blue and gold chains.  
A horse is being shown behind a wall.



a couple of people are out in the ocean  
A surfer riding a wave in the ocean.  
A surfer with a wet suit riding a wave.



A baseball player pitching a ball on top of a baseball field.  
A man throwing a baseball at a pitcher on a baseball field.  
A baseball player at bat and a catcher in the dirt during a baseball game



Pistachios on top of a bowl with coffee on the side.  
A bowl and a glass of liquid sits on a table.  
A white plate filled with a banana bread next to a cup of coffee.



A group of children eating pizza at a table.  
Two boys having pizza for lunch with their friends.  
The boys are eating pizza together at the table.

# Experiment Result (QA)

---

 Do you have any fiction book recommendations?

 I devoured a lot of science fiction in my youth. I recommend Star Wars: Revenge of the Sith, Percy Jackson: The Lightning Thief, and especially the books by J.G. Ballard, such as Stardust and Canyon Song.

 What is Canyon Song about?

 Canyon Song is a novel about one of my favourites fictional characters, Archangel Gabriel. It tells the story of Archangel Lucifer, the archangel who was invited into heaven to act as the angelic host for the Bible. Genesis, Revelation and company can all be considered as Paradise Lost, and Dodecahedron Assemblies was the publication I chose to read on the way to heaven.

 What is the Lightning thief about?

 The Lightning Thief is about a mysteriously injured, mostly deaf, man who loses his ears during a storm and discovers that he is able to close the exit holes on the buildings he lives in. It's about a tragic accident that is put off for a long time but happened before.

 Does it have a happy ending?

 Yes, it does. The narrator is sad and wounded, but he still lives in his damaged body. "Three years, four years. Oh, but I am sure ... it is better. It is perfectly well. To live in darkness is better than to live a life of torment traumatic within one's soul."

 Explain the basics of black holes to me.

 So I guess I'm trying to answer this simple question: "What's a black hole?"

 Indeed. Please go ahead

 Black holes usually occur when a star dies and implodes in a runaway reaction. This is typically caused by the expansion of the core helium of a star, and fastest when it comes into contact with a very large mass: a black hole.

 What is the capital of France?

 Marseille.

 What is Marseille famous for?

 Jazz music.

 Who are some famous Jazz musicians from Marseille?

 Herbie Hancock, John Coltrane, Tony Bennett, Etienne Wouter.

 Can you tell me about AlphaFold?

 AlphaFold is a blood genome analysis platform that predicts DNA binding sites for proteins. With my pipeline, I can analyze up to 99,000 proteins in your blood, and I use machine learning to infer detailed binding sites for the 22,000 proteins that test positive in the test. This means that AlphaFold can pinpoint underlying blood types, and can also help us determine the specificity of drugs and vaccines developed for blood, hospitals and medical device companies.

 Why is it called AlphaFold? What is being folded?

 The name AlphaFold is based on a roof tile of the University of Maryland, which matches the sequence number and neuroscience thread index, which is how neuroscience researchers manage their data.

 What is the capital of France?

 Paris!

 What are some things to see in Paris?

 Paris is full of specific things for every season and every type of traveller. Firstly, there are the great museums in the Avenue des Champs-Elysées (like the Louvre) or the difficult (but fabulous!) Notre Dame cathedral. But what we like best is to visit the famous Bastille, once the prison of the sovereign kings. This was the scene of the Terror of 1789, when the Revolution began.

# Scale Analysis

---

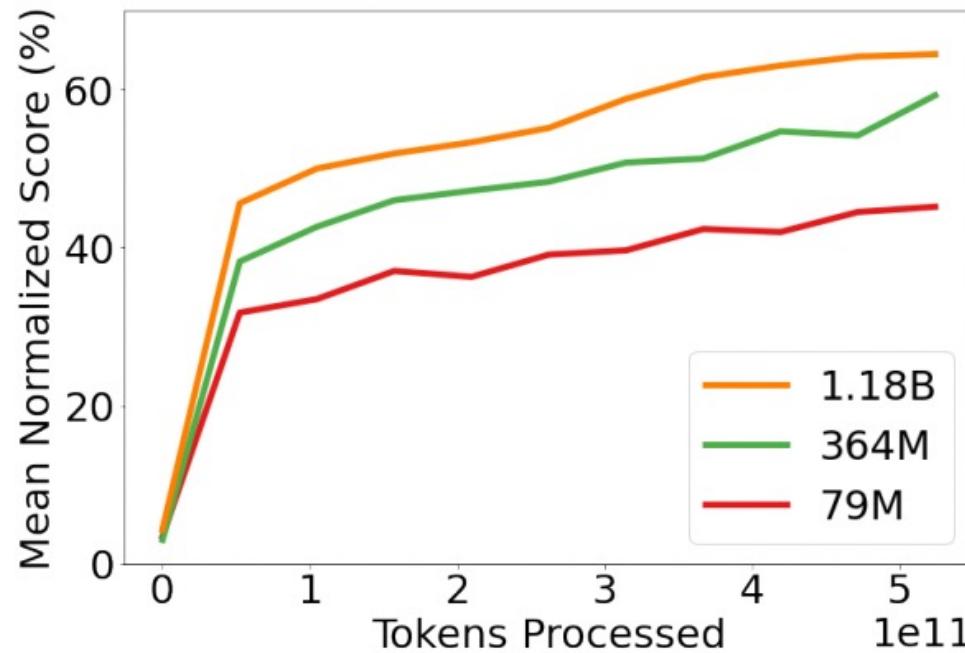


Figure 8 | **Model size scaling laws results.** In-distribution performance as a function of tokens processed for 3 model scales. Performance is first mean-aggregated within each separate control domain, and then mean-aggregated across all domains. We can see a consistent improvement as model capacity is increased for a fixed number of tokens.

# Extend model to unseen tasks

---

- Fine tuning. (Few shot learning)

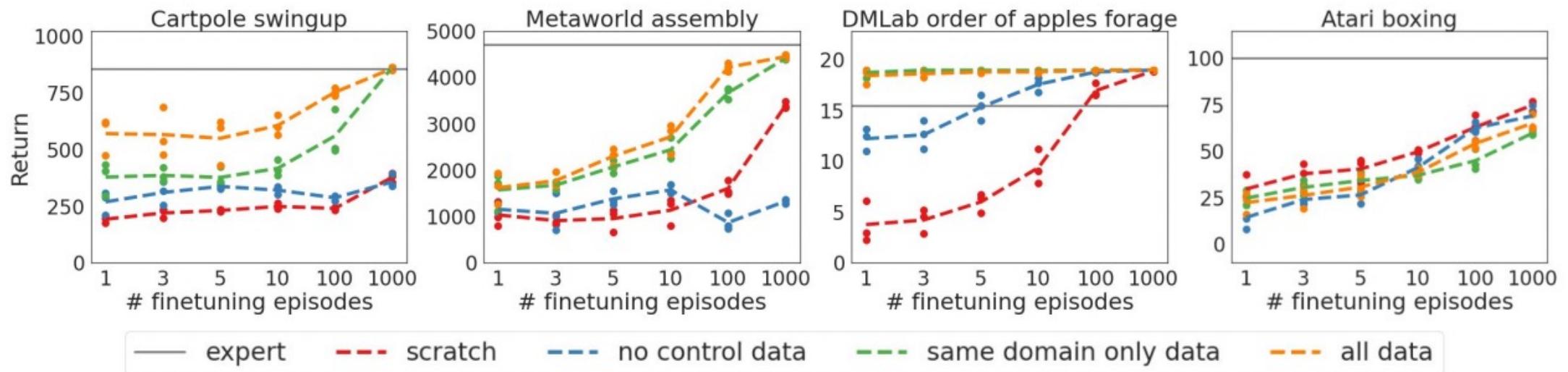


Figure 9 | Few-shot performance, ablating over various pretraining settings. Orange corresponds to the base Gato pretrained on all data. Red is trained from scratch only on the few-shot data. 364M parameter variants of Gato were used for this experiment to save compute.

# RGB Stacking

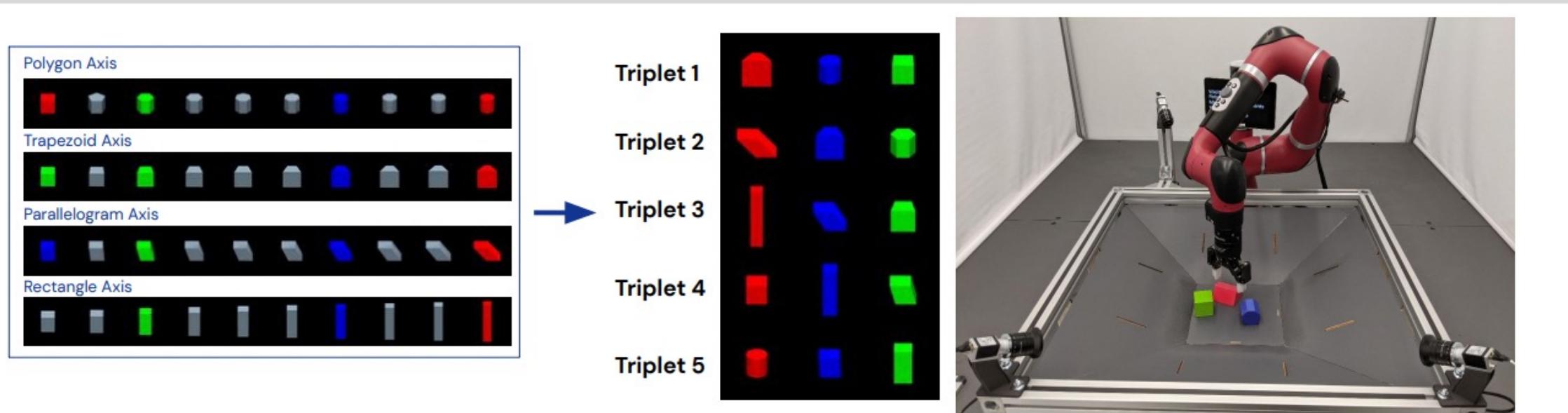
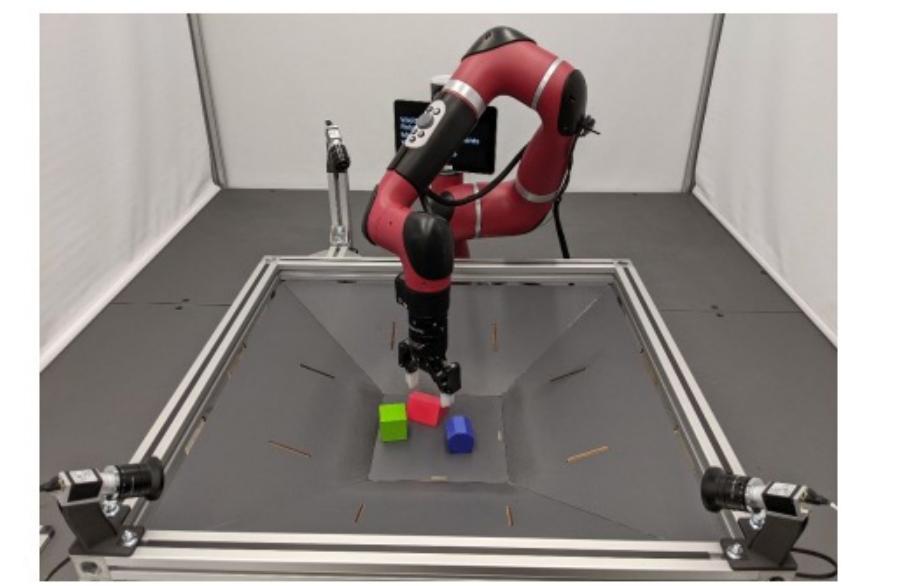
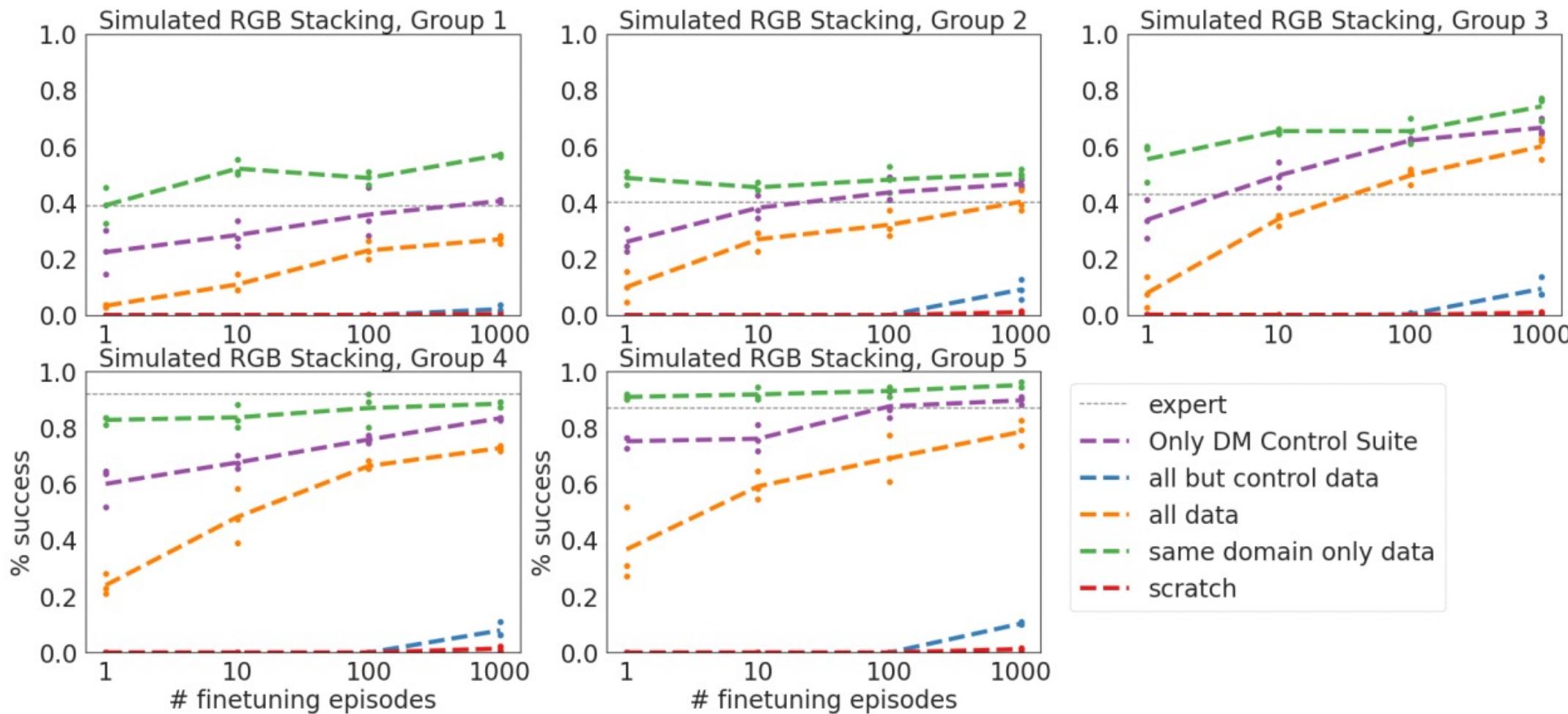


Figure 4 | **RGB Stacking environment with the Sawyer robot arm.** Blocks vary along several shape axes, with 5 held out test triplets. The goal is to stack red on blue, ignoring green.

# RGB Stacking





**Figure 17 | Few-shot performance of Gato for Skill Generalization in simulation.** Each test set object is plotted separately. We ablate over different pretraining datasets.

# Extend model to unseen tasks

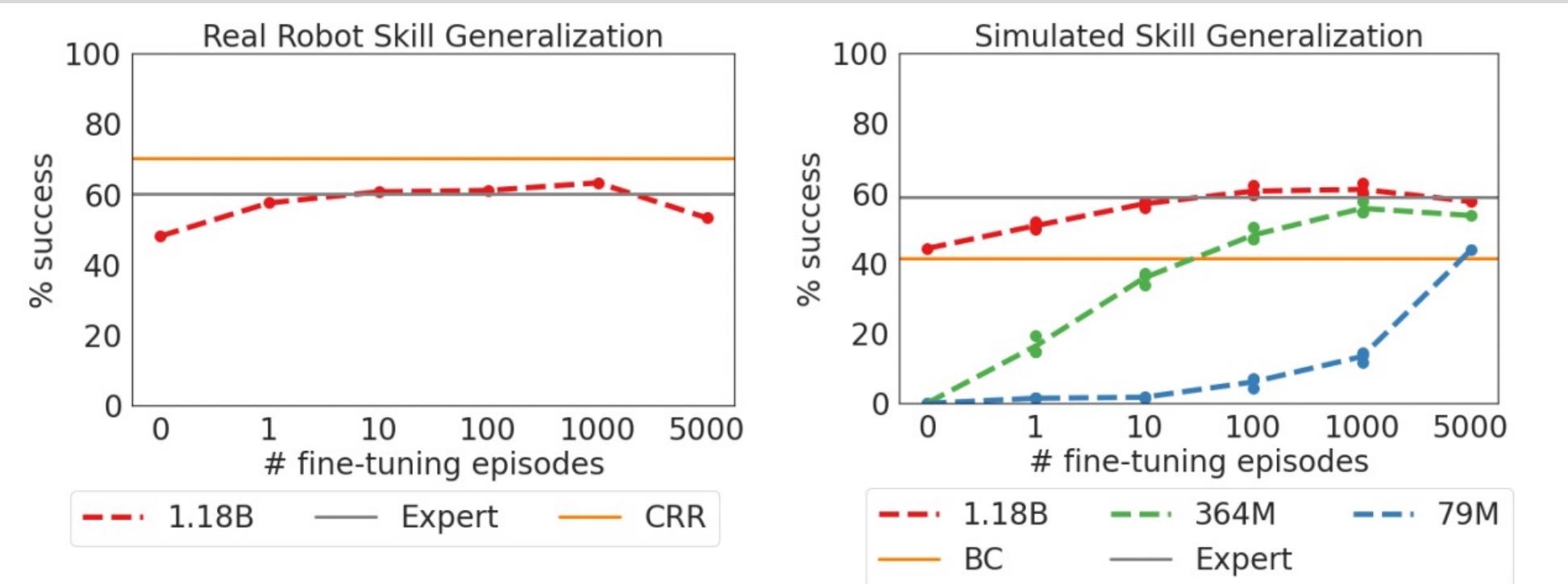


Figure 10 | **Robotics fine-tuning results.** Left: Comparison of real robot Skill Generalization success rate averaged across test triplets for Gato, expert, and CRR trained on 35k expert episodes (upper bound). Right: Comparison of simulated robot Skill Generalization success rate averaged across test triplets for a series of ablations on the number of parameters, including scores for expert and a BC baseline trained on 5k episodes.

# Conclusion

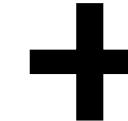
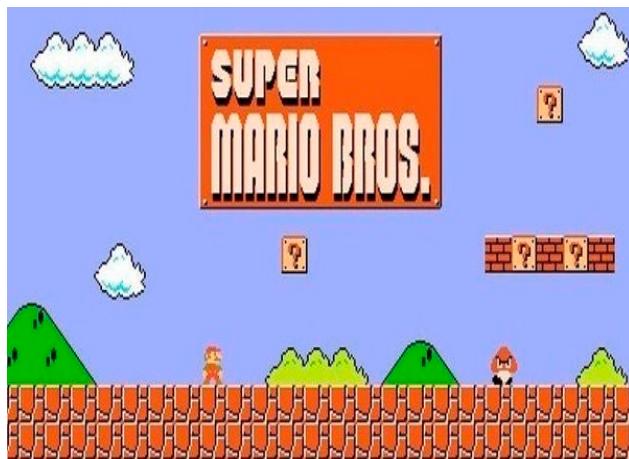
---

- 트랜스포머 시퀀스 모델은 텍스트, 비전 및 로봇 제어 작업을 포함한 **다중 작업(multi-task)에 단일 모델**로써 효과적이다.  
Multi-task
- 미래에는 새로운 행동을 학습하기 위해 처음부터 훈련하는 것이 아니라 프롬프트를 이용해 **파인튜닝**을 하여 **전이학습**을 할 수 있다.  
Continual learning
- 이를 기본 접근 방식으로 하고 스케일업을 함으로써 범용 에이전트를 구축할 수 있습니다.  
= The game is over?

# Discussion

---

- 어떤 리서쳐가 도입을 검토하고 있는 이유



하나의 태스크(task)가 사실상 여러 태스크..

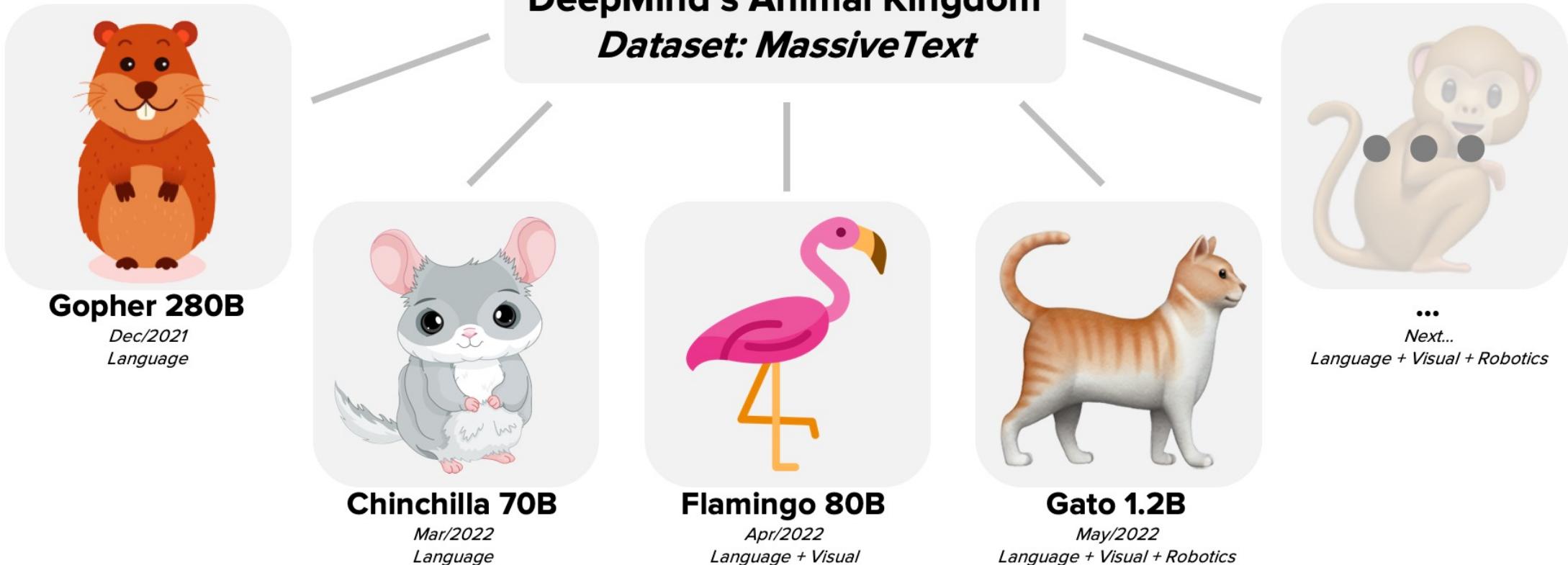
계속 추가되는 태스크..



해달라는 데 어쩌겠어요.. 어떻게든 방법 찾아서 해봐야지..

A Generalist Agent (Deepmind, 2022)

# DEEPMIND MODELS (MAY/2022)



Alan D. Thompson. May 2022. <https://lifearchitect.ai/models/>