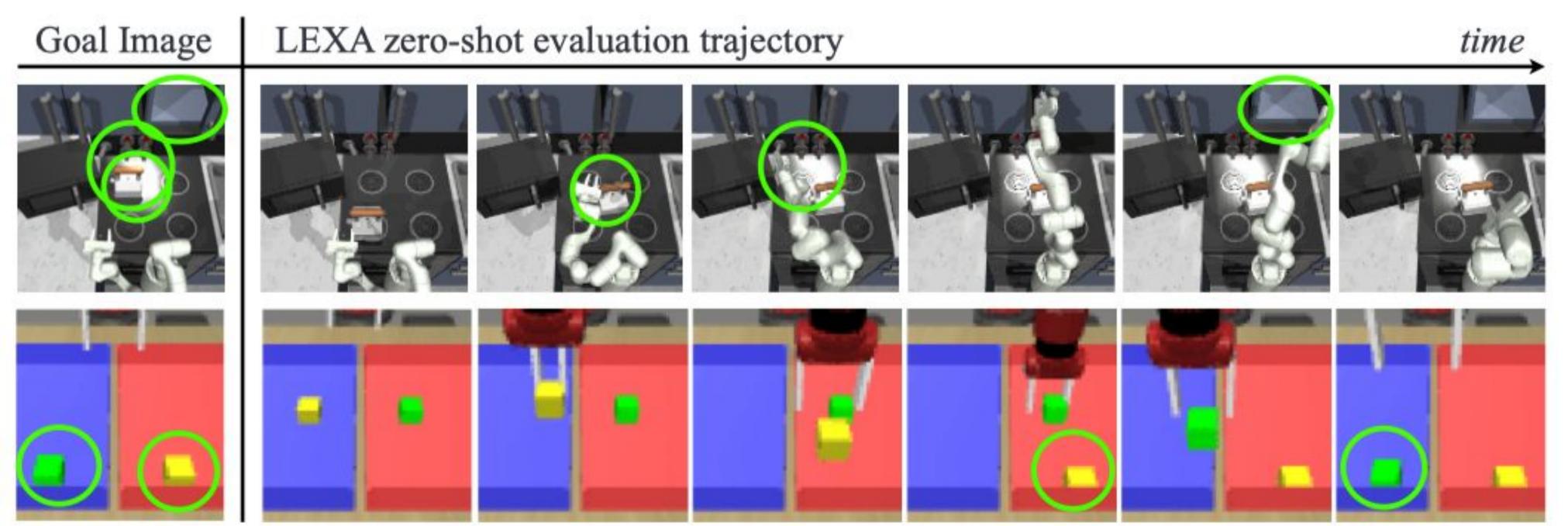


# **Discovering and Achieving Goals via World Models** Russell Mendonca<sup>1</sup>\*, Oleh Rybkin<sup>2</sup>\*, Kostas Daniilidis<sup>2</sup>, Danijar Hafner<sup>3</sup>, Deepak Pathak<sup>1</sup>

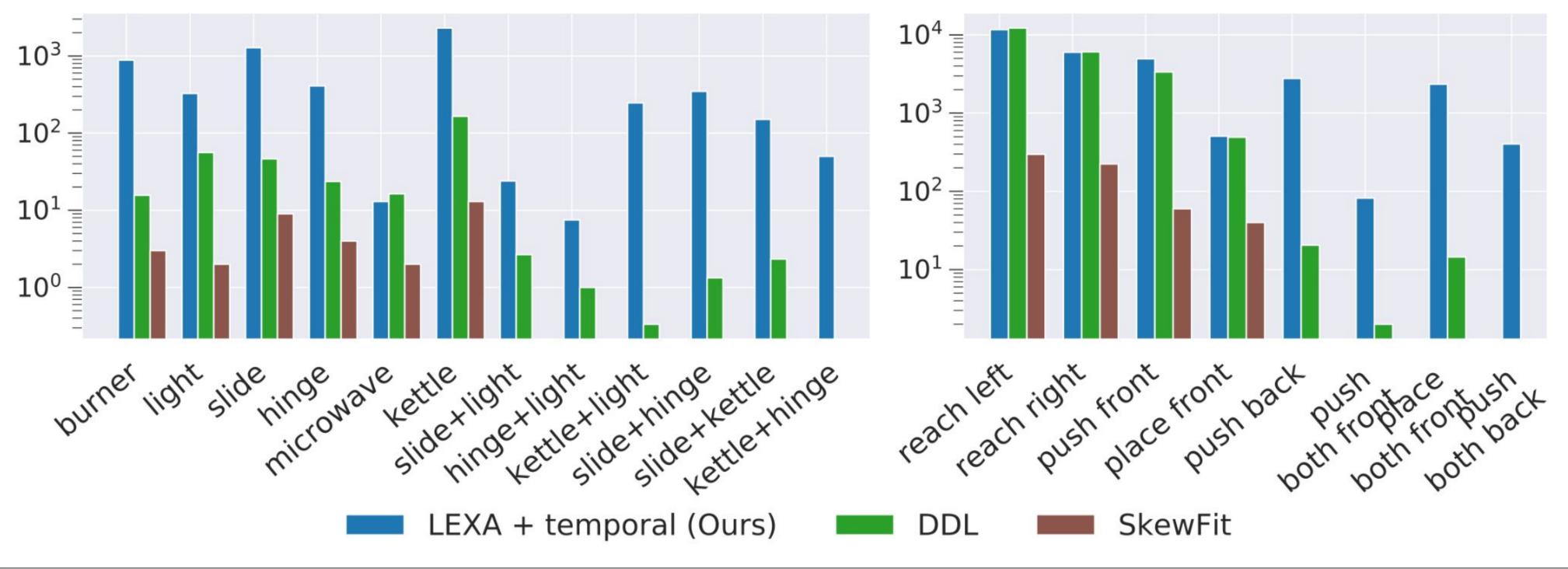
1 Introducing LEXA Learns a world model without any supervision, trains two policies (explorer and achiever) in imagination. The explorer finds new images via model disagreement and the achiever learns to reliably reach them. Once trained, the achiever reaches user-specified goals 3 zero-shot without further training at test time. **Unsupervised Interaction** Achieve User Goal Environment World Model Explorer

# 4 Qualitative Performance

LEXA moves multiple objects to solve a single goal image

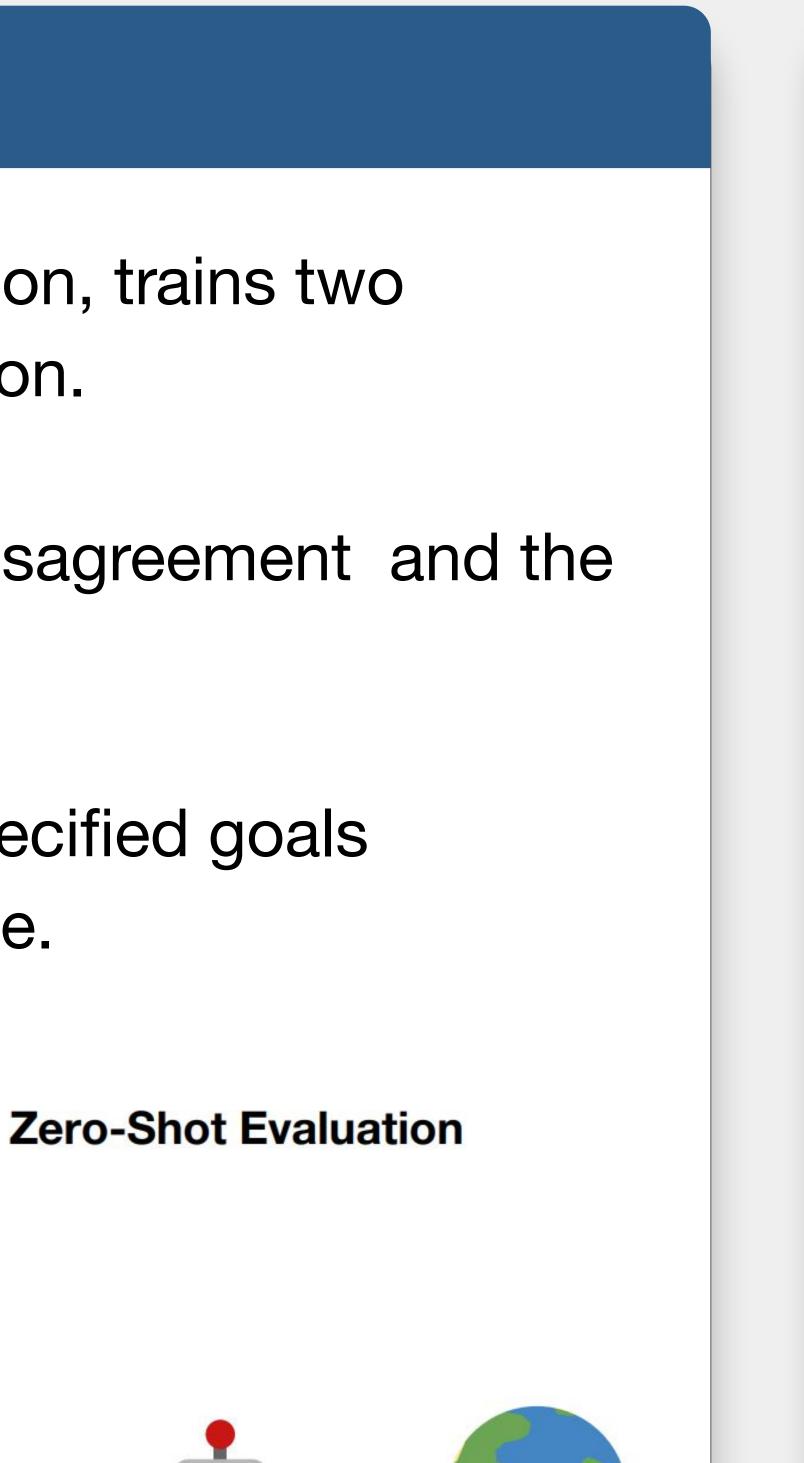


Coincidental goal success during unsupervised exploration



\* equal contribution <sup>1</sup>Carnegie Mellon University <sup>2</sup>University of Pennsylvania

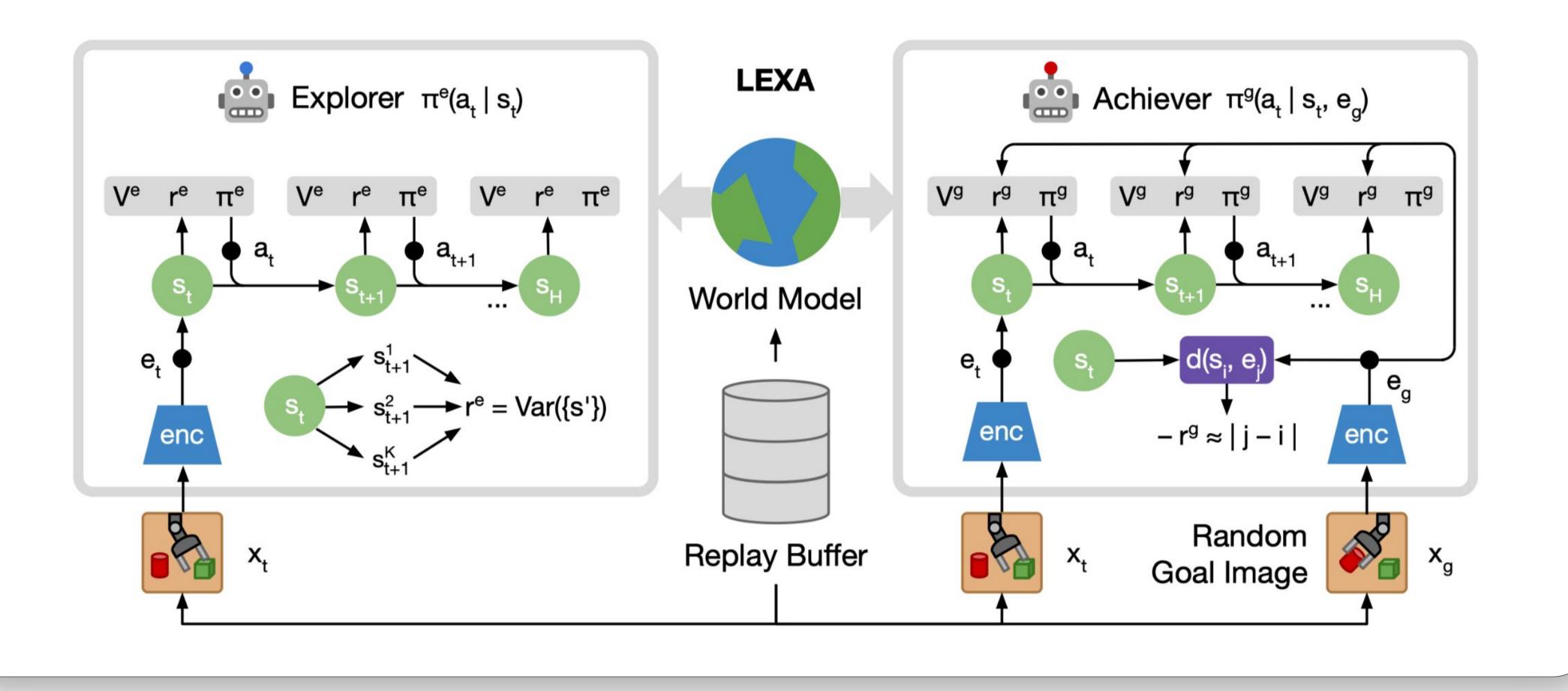
(3)



Environment

## 2 Algorithm Details

- model to maximize the disagreement objective
- 2



## 5 Benchmark Scores

## Prior benchmarks evaluation

			Task	LEXA	DISCERN
Method	Pusher	Pickup	Cup	84.0%	76.5%
RIG [33]	7.7cm	3.7cm	Cartpole	35.9%	21.3%
RIG + HER [2]	7.5cm	3.5cm	Finger	40.9%	21.8%
Skew-Fit [38]	4.9cm	1.8cm	Pendulum	79.1%	75.7%
LEXA + Temporal	<b>2.3cm</b>	<b>1.4cm</b>	Pointmass	83.2%	49.6%
			Reacher	100.0%	87.1%

## New benchmarks - Final goal success percentages

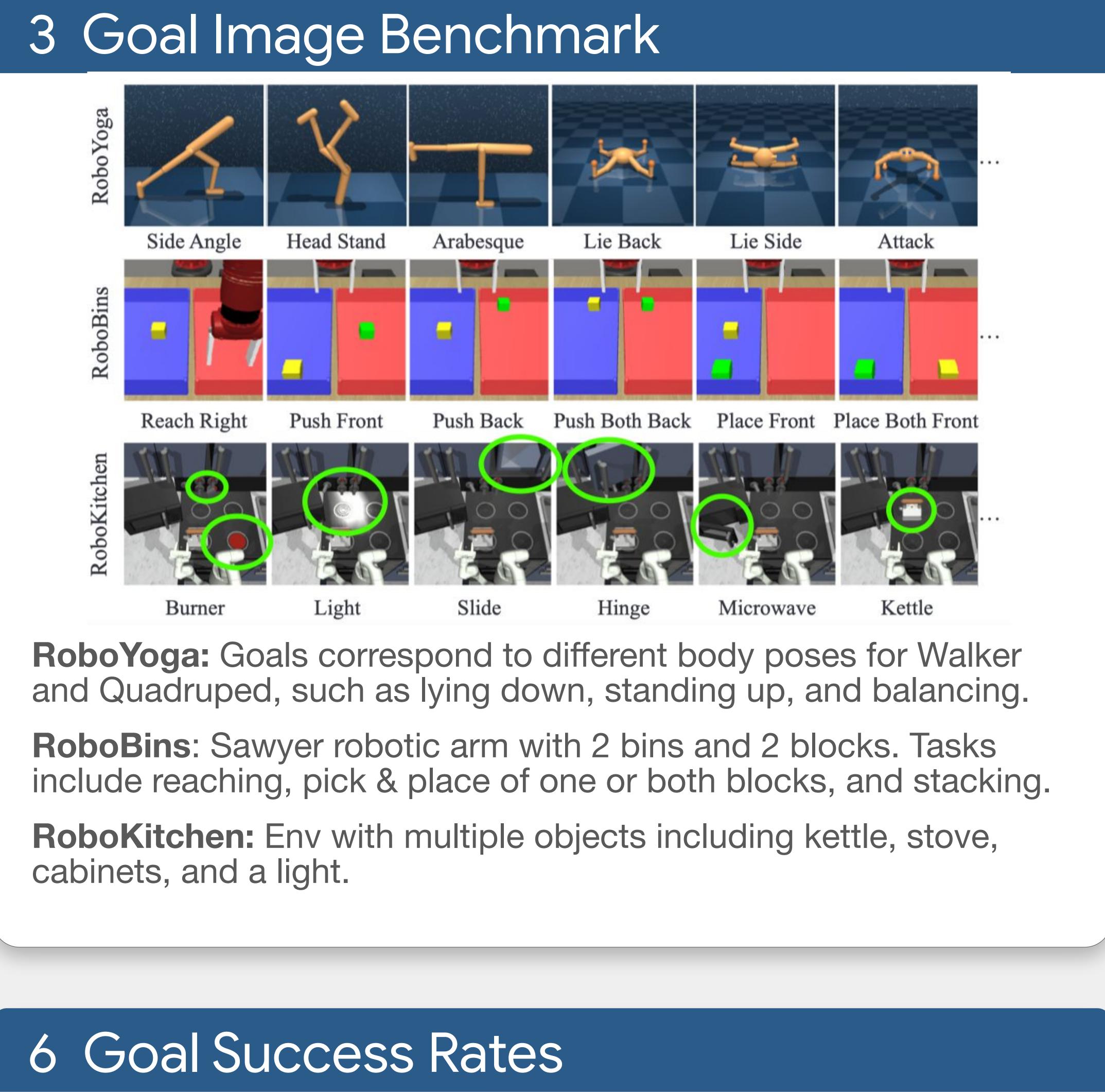
Method	Kitchen	RoboBins	Quadruped	Walker
DDL	0.00	35.42	22.50	40.00
DIAYN	0.00	13.69	13.81	0.28
GCSL	0.00	7.94	15.83	1.11
SkewFit	0.23	15.77	5.52	0.01
LEXA + Temporal (Ours) LEXA + Cosine (Ours)	<b>37.50</b> 6.02	<b>69.44</b> 45.83	31.39 <b>56.11</b>	36.72 <b>73.06</b>

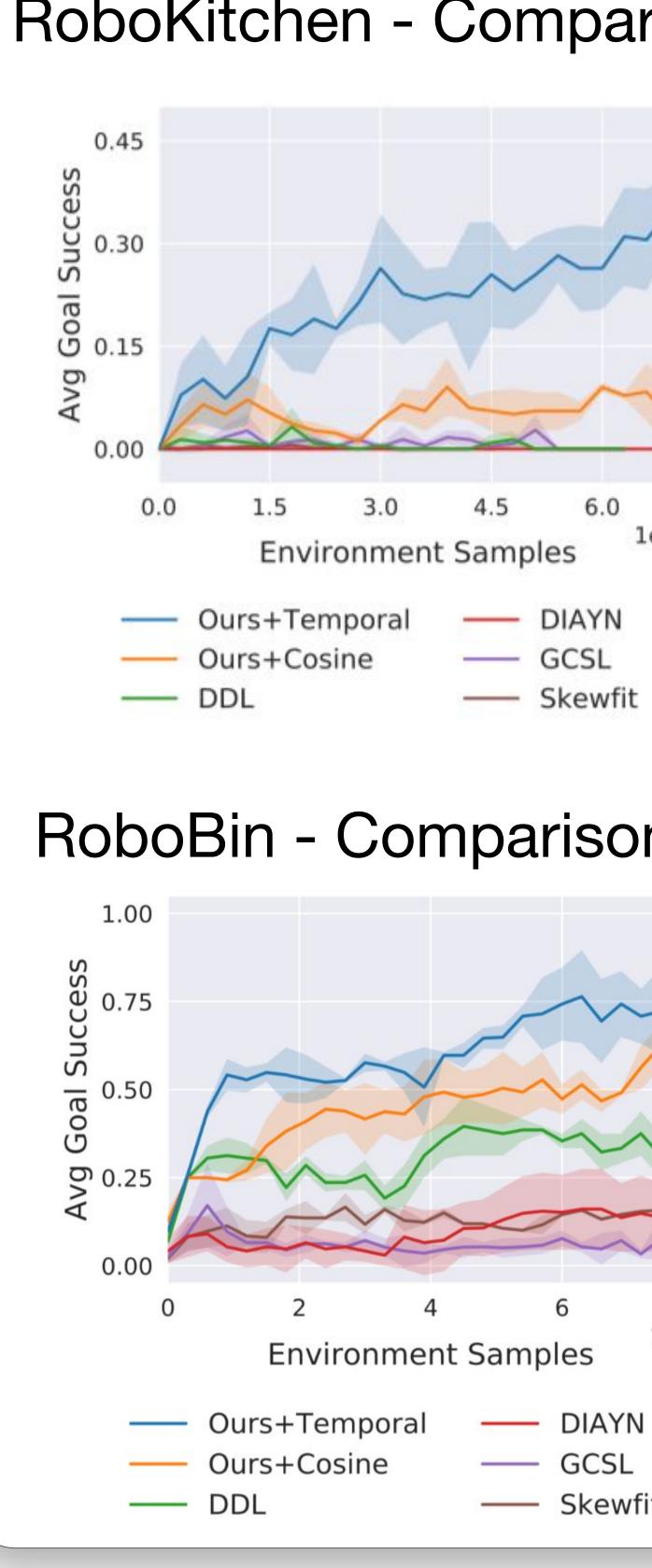
# <sup>3</sup> University of Toronto

Explorer is trained on imagined latent state rollouts of the world

Achiever policy is conditioned on goals (sampled from collected data) and trained in imagination to minimize a distance function.

Trains a temporal distance function, using the imagined rollouts of the achiever to predict number of time steps between states.





## Website with code and videos: https://orybkin.github.io/lexa/

GCSL





### RoboKitchen - Comparisons for per-task and averaged performance

	Ours+ Temporal	1	0.83	0.39	0.67	0.5	0.28	0.28	0.17	0.06	0	0	0
~~~	Ours+ Cosine	0.5	0.08	0.5	0	0	0	0	0	0	0	0	0
	DDL	0	0	0	0	0	0	0	0	0	0	0	0
	DIAYN	0	0	0	0	0	0	0	0	0	0	0	0
4.5 6.0 Samples	GCSL	0.05	0.24	0	0	0	0.05	0	0	0	0	0	0
DIAYN GCSL Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Skewfit Ske													
	Ours+ Temporal			0.94	1	(	0.89	0.56	0	.94	0.44	0.33	5
m	Ours+ Cosine			1	1		1	0.5	C	).5	0	0	
	DDL	1		1	0.83	3	0	0	0	.17	0	0	
	DIAYN	0.6	52	0	0.6	7	0	0		0	0	0	

SkewFit 1e6