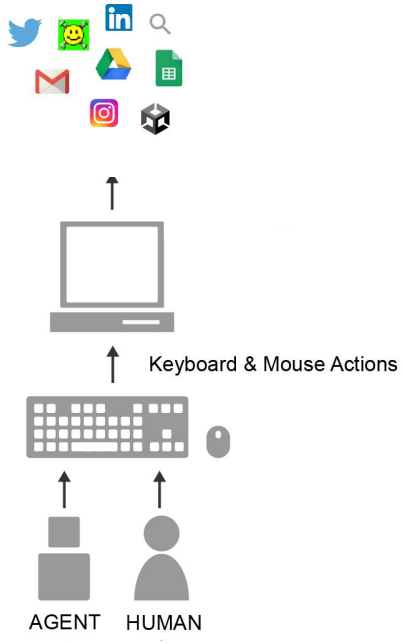
DeepMind

A Data-Driven Approach for Learning to Control Computers

Peter Humphreys, David Raposo, Toby Pohlen, Gregory Thornton, Rachita Chhaparia, Alistair Muldal, Josh Abramson, Petko Georgiev, Adam Santoro, Timothy Lillicrap

Introduction

It would be useful for machines to use computers as humans do, so that they can aid us in everyday tasks.

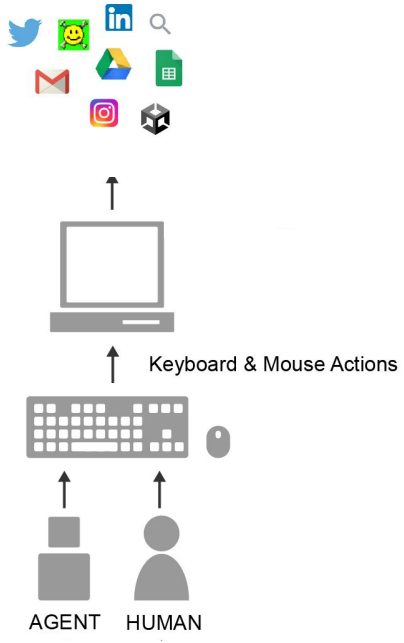


Introduction

It would be useful for machines to use computers as humans do, so that they can aid us in everyday tasks.

We investigated computer control using keyboard and mouse, with goals specified via natural language.

- Supports transfer & generalisation across tasks.
- Leverage human experience and feedback.



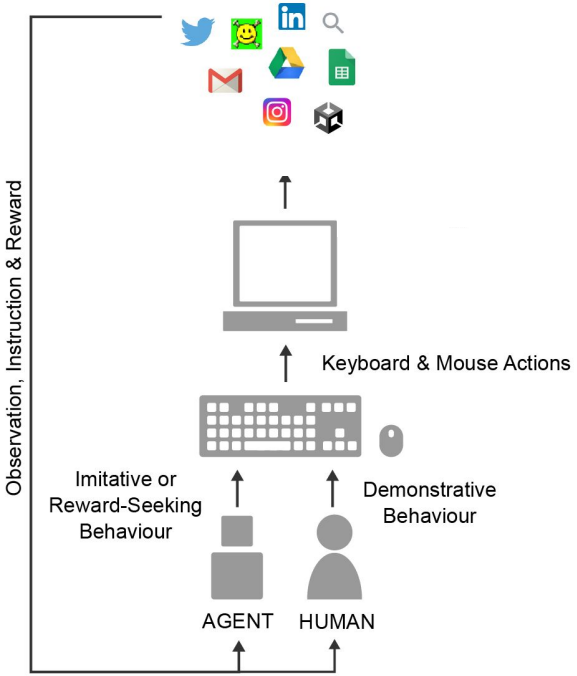
Introduction

It would be useful for machines to use computers as humans do, so that they can aid us in everyday tasks.

We investigated computer control using keyboard and mouse, with goals specified via natural language.

- Supports transfer & generalisation across tasks.
- Leverage human experience and feedback.

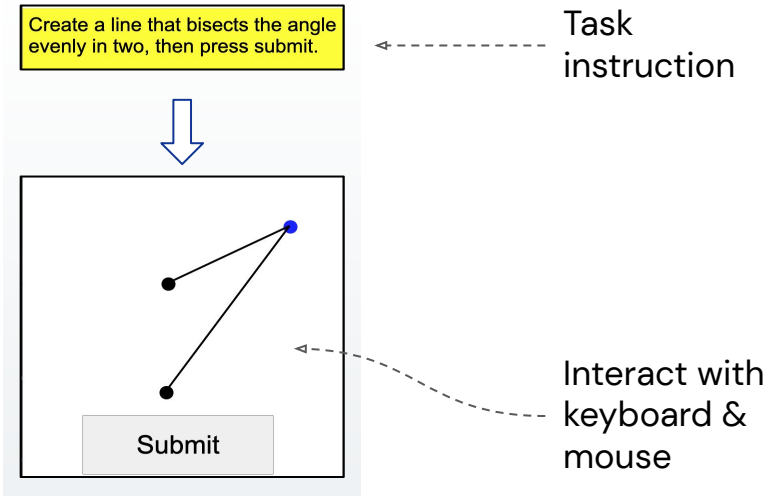
We focus on developing a scalable method centered on **reinforcement learning** combined with **behavioural priors** informed by actual human-computer interactions.



MiniWob++ Task Suite

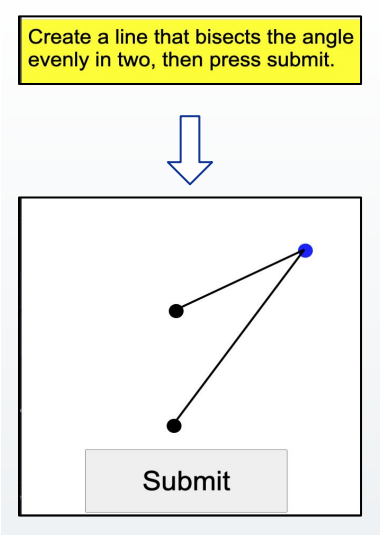
A suite of **104 web-browser based tasks** introduced in Liu et al. (2018) (an extension of the earlier MiniWob task suite (Shi et al., 2017)).

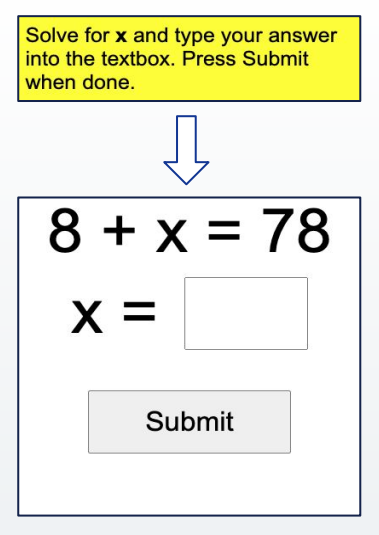
Programmatic rewards are available for each task.



MiniWob++ Task Suite

Tasks range from simple button clicking to complex form-filling.

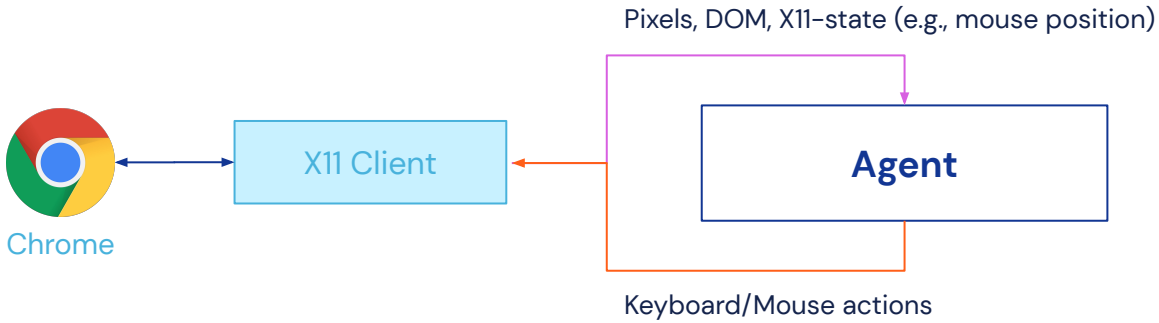




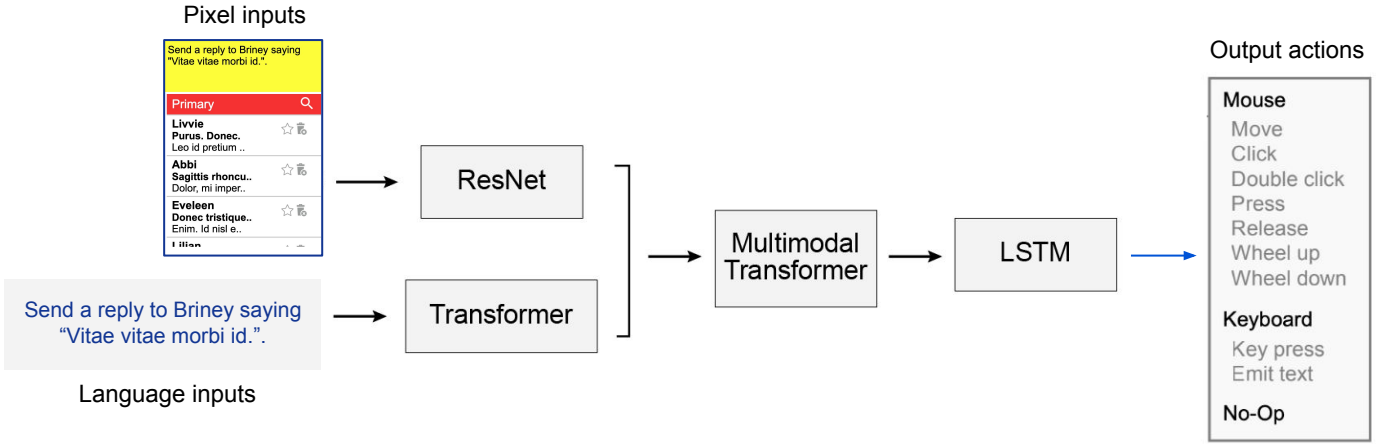
Interface & human data collection

- Our environment connects directly to an X11 server to input mouse & keyboard controls and retrieve observations. This allows agents to use the **same controls as humans** and **minimises domain shift**.

- This interface simplifies recording demonstration data at scale. For this study, we collected over **2.4 million demonstrations** of the MiniWob++ tasks (~1000 times more than previous studies).



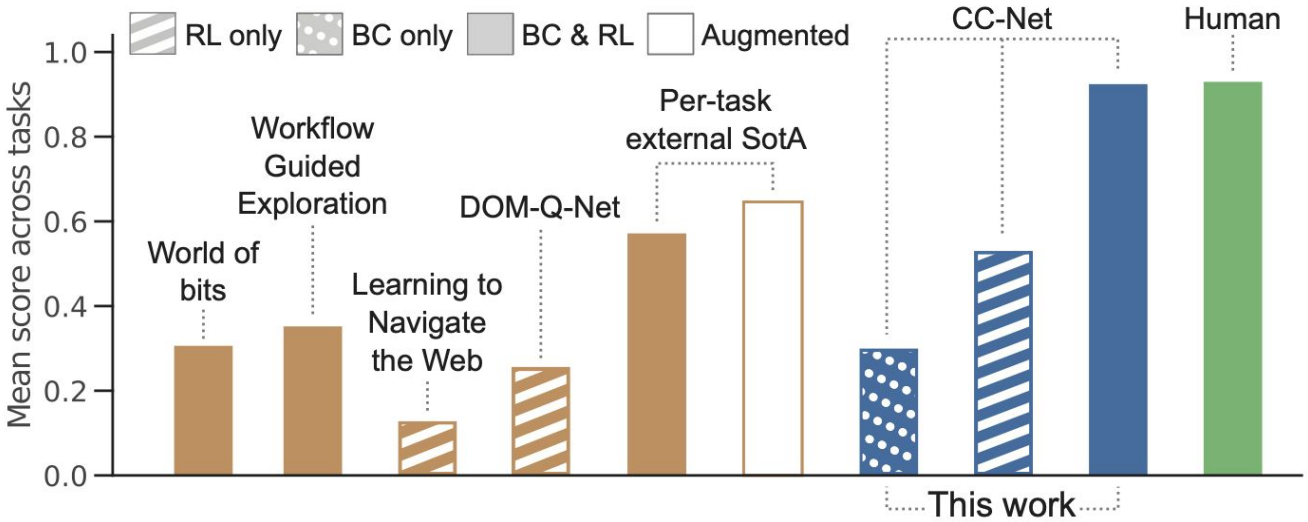
MiniWob Agent



Trained with behavioural cloning (BC) on human data + VMPO (RL) on agent experience.

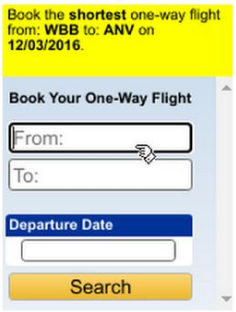
$$\mathcal{L}(\theta) = \mathbb{E}_{\tau_{\text{EXPERT}}}[-\ln \pi_\theta(\mathbf{a} \mid \mathbf{o})] + \mathbb{E}_{\tau_{\text{AGENT}}}[\mathcal{L}^{\text{VMPO}}(\mathbf{a}, \mathbf{o})]$$

Results

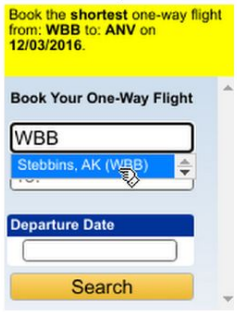

RL only
BC only
BC & RL
Augmented
Mean score across tasks
1.0
0.8
0.6
0.4
0.2
0.0
World of bits
Workflow Guided Exploration
Learning to Navigate the Web
DOM-Q-Net
Per-task external SotA
CC-Net
Human
This work

Results

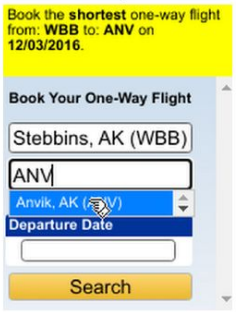
Step 3



Step 6

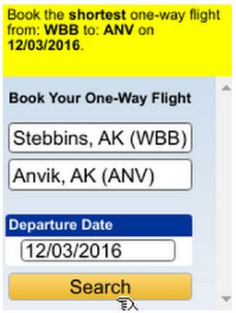


Step 9

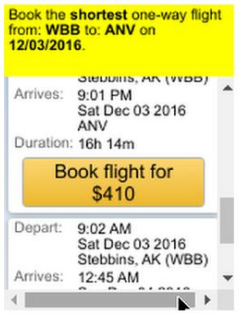


Step 16

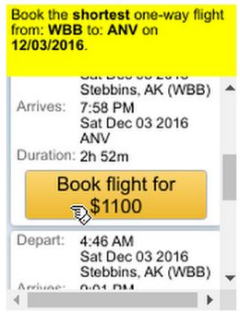
Step 17



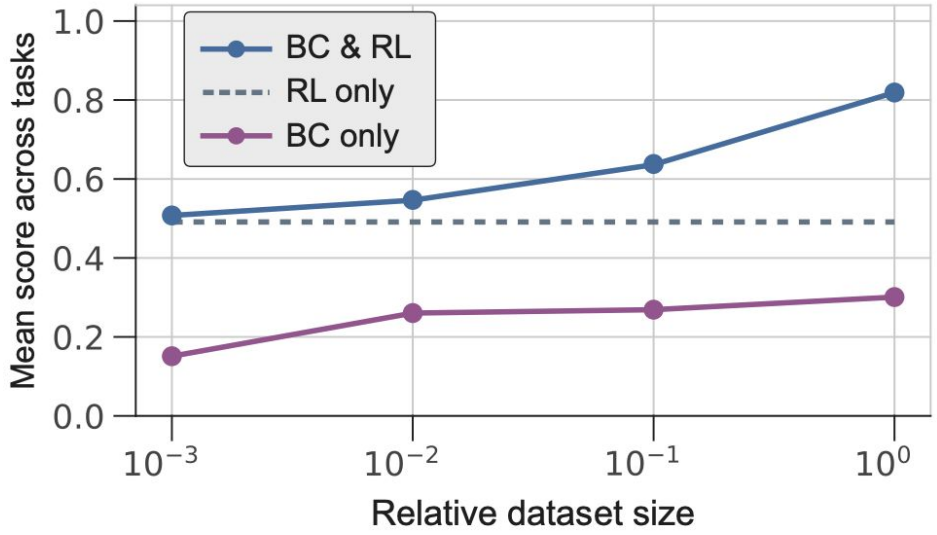
Step 26



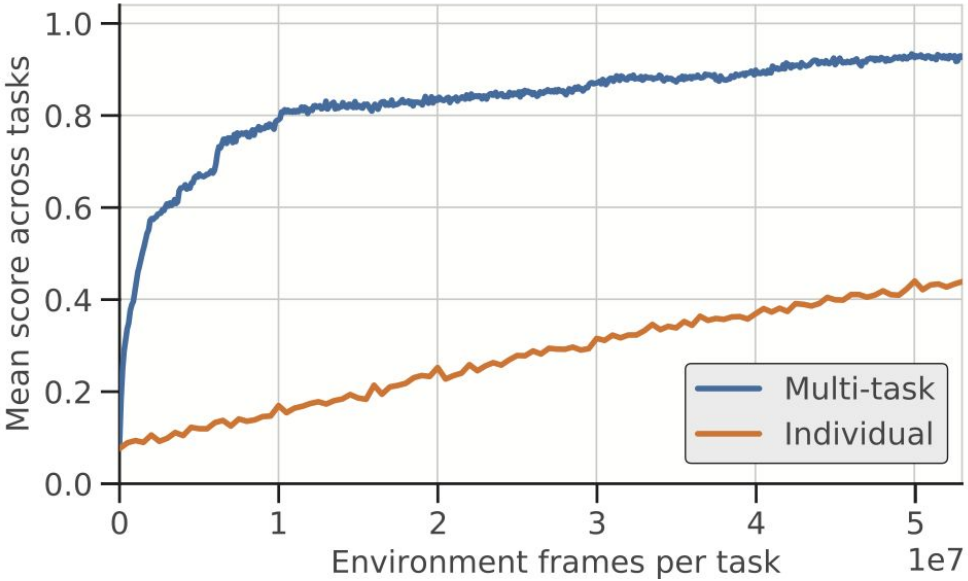
Step 34



Data scale was important



Multi-task training was beneficial



DeepMind

Thank you!

See us at poster #836