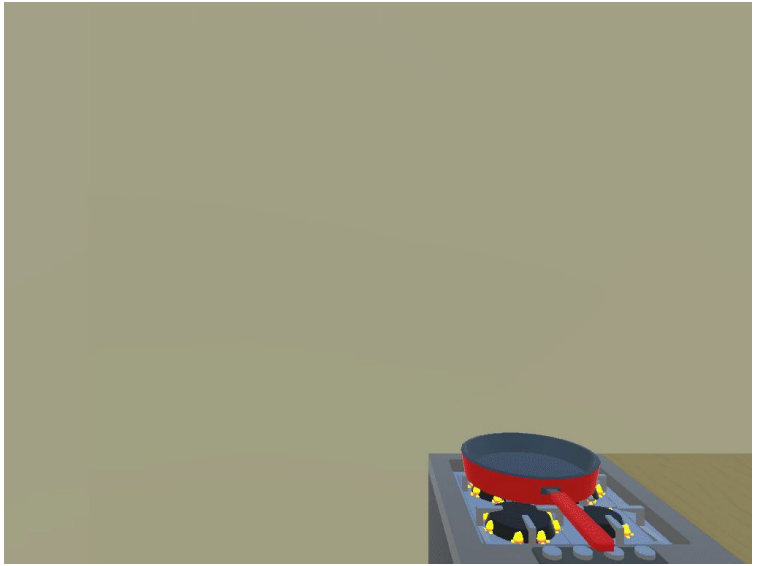
Today’s Focus:
Model Reasoning with Sub-Goal Annotations

Ryan Faulkner February 2025

Motivation for Sub-Goals

In our current agent setup we fine tune a language model to produce low level actions directly from an input frame and a high level task instruction.

Cook the burger patty.



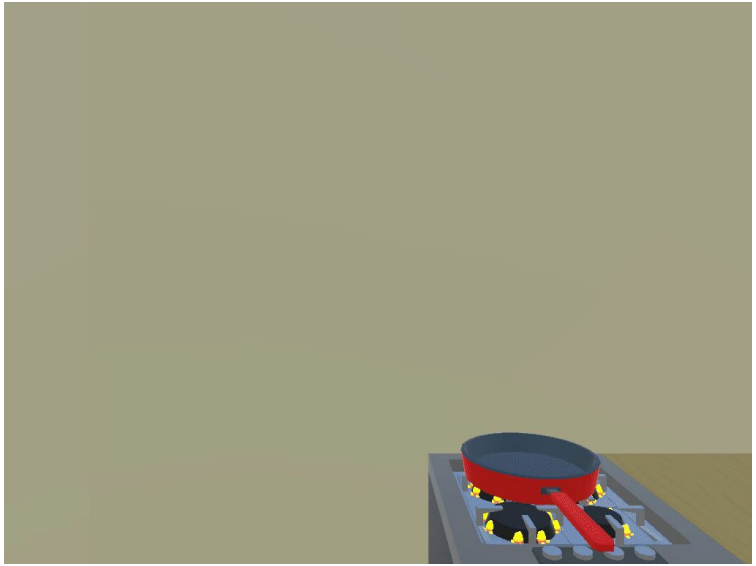
```
KeyD,KeyD,KeyD,KeyD,MouseRight=8+MouseDown=5,MouseRig
ht=11+MouseDown=3,MouseRight=7+MouseDown=2,MouseRight
=2+MouseDown=1,MouseRight=3+MouseDown=1+ClickLeft,Key
S+ClickLeft,KeyS+ClickLeft,KeyS+ClickLeft,KeyS+ClickL
eft,KeyA+ClickLeft,KeyA+ClickLeft,KeyA+ClickLeft,KeyA
+ClickLeft,KeyW+ClickLeft,KeyW+ClickLeft,KeyW+ClickLe
ft,KeyW+ClickLeft,MouseRight=11+MouseDown=3+ClickLeft
,MouseRight=7+MouseDown=2+ClickLeft,MouseRight=2+Mous
eDown=1+ClickLeft,MouseRight=7+MouseDown=2,ClickLeft,
ClickLeft,MouseLeft=2
```

Motivation for Sub-Goals

In our current agent setup we fine tune a language model to produce low level actions directly from an input frame and a high level task instruction.

Action sequences can be very verbose, lengthy, and redundant, even in the interleaved mode.

Cook the burger patty.



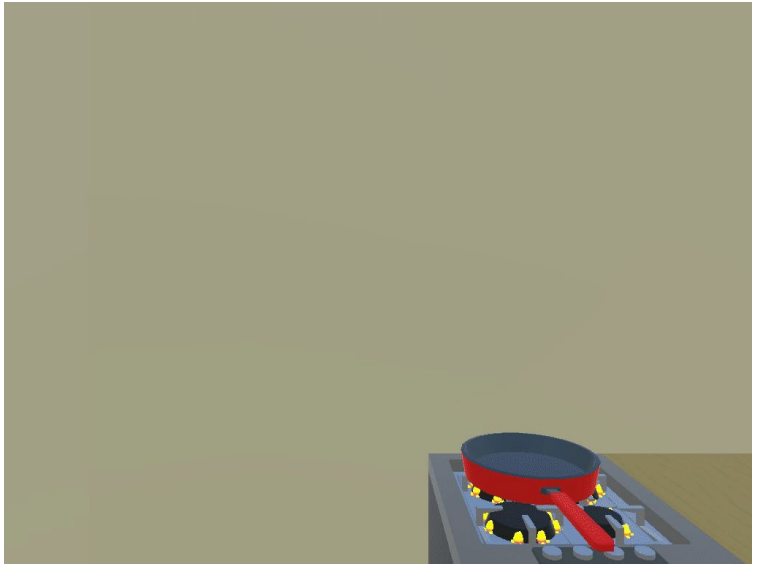
Motivation for Sub-Goals

In our current agent setup we fine tune a language model to produce low level actions directly from an input frame and a high level task instruction.

Action sequences can be very verbose, lengthy, and redundant, even in the interleaved mode.

Therefore, it may be helpful to break a task up into sub-tasks with more specific instruction grounded in the scene and KBM actions.

Cook the burger patty.



"The player moves right with keypad and left clicks to hold patty."
"Then navigates back, left and forward with the keypad to the stove."
"Finally places the patty into the pan by moving the mouse and releasing the mouse button."

Motivation for Sub-Goals

In our current agent setup we fine tune a language model to produce low level actions directly from an input frame and a high level task instruction.

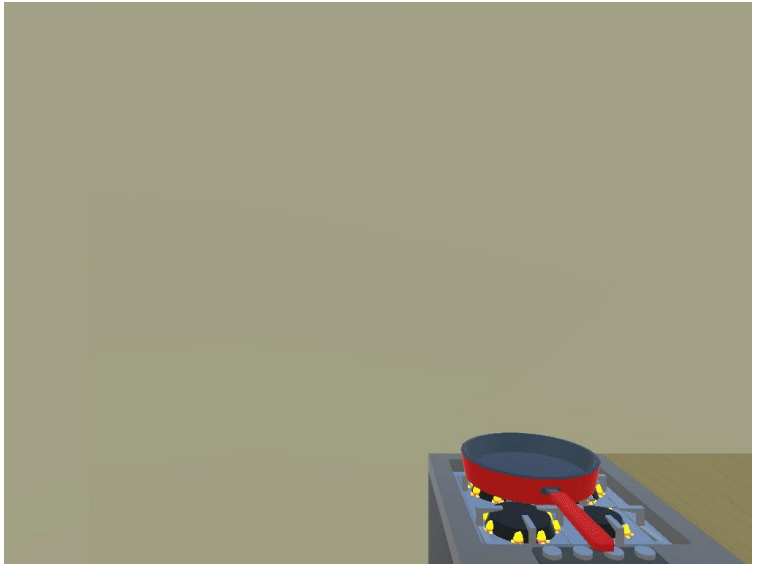
Action sequences can be very verbose, lengthy, and redundant, even in the interleaved mode.

Therefore, it may be helpful to break a task up into sub-tasks with more specific instruction grounded in the scene and KBM actions.

This could:

- **Improve acting by reasoning through explicit sub-goals that use hindsight.**
- **Provide a more flexible interface to prompt the model toward desired behaviours. [+interpretability]**
- **Aid in solving more complex tasks by planning through reasoning**
- **Help us towards retaining language capability.**

Cook the burger patty.



"The player moves right with keypad and left clicks to hold patty."
"Then navigates back, left and forward with the keypad to the stove."
"Finally places the patty into the pan by moving the mouse and releasing the mouse button."

Instructions often under-specify the span behaviour, where behaviour can be split into a series of smaller behaviour sub-goals.

"run towards straight to reach the branch"

"recharge the hazard protection using the sodium in the exosuit menu"

Making KBM Action-Goals

- Using an interleaved setup, **pick key frames**: fixed, metrics, **LLM**, etc.
- The annotations will depend on how we decide to partition the span
- Most effective partitions will be consistent with the behavioural semantics


Actions

Actions
...

Making KBM Action-Goals

- For each segment prompt an annotation model.
- Sub-goals are from **hindsight** (ie. grounded in the actual frames and actions)

<explicit actions sequence>

You are watching a player play a video game ...

The player's movements are controlled by keyboard and mouse ...

Given the ground truth actions and video please describe the actions in natural language.

Some examples include:
[example 1] [example 2] ...

Annotations

Actions

Actions

...

Making KBM Action-Goals

Context info from the key frame can be referenced

Attempt to summarise actions (reduce repetition)

Currently works on the span level resolution.

- It may be useful instead to do this on a broader scale

footpath

wall

red box

Annotations

Actions

Actions

...

Condition Actions on Action-Goals

Write results to a datatable (annotations, key frame indices, collection sequence ID)

Run follow up pipeline that generates the data with interleaved annotations

Generated Spans

Generated Data - NMS

200 Ground Truth Actions

Datatable record
S2 Sequence

Instruction:

"recharge the hazard protection using the sodium in the exosuit menu"

Annotations:

The player clicks on the Portable Refiner multiple times to select and interact with it.||", "The player moves the mouse across the screen to select different items in the inventory.||", "The player moves the mouse over the Condensed Carbon item several times, then clicks and drags the mouse to select multiple units.||", ... more annotations

Generated Data - Valheim

208 Ground Truth Actions

Datatable record
S2 Sequence

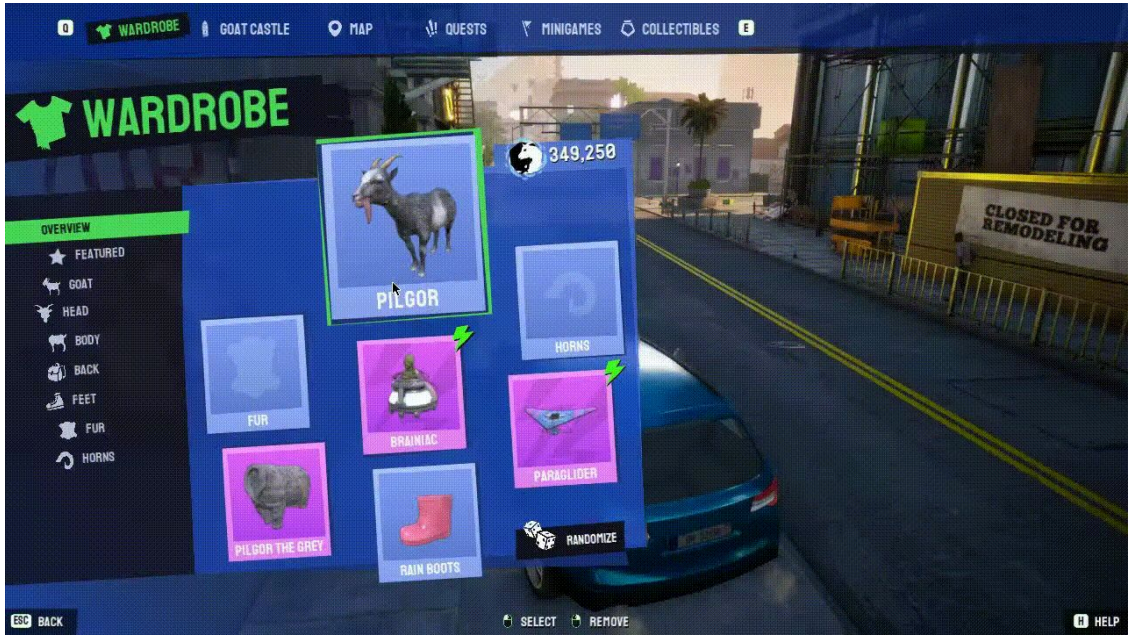
Instruction:

"run towards straight to reach the branch"

Annotations:

"The player moves the mouse to look slightly to the right. The player then moves the mouse further to the right and slightly down.||" .... "The player moves the mouse to look slightly to the right while holding down the W key and left shift key to move forward and run.||" ...

Generated Data - Factory



33 Ground Truth Actions

Datatable record
S2 Sequence

Instruction:

"open the map menu."

Annotations:

"The user moves the mouse to select the Pilgor the Grey item, then moves the mouse to select the Dain Boots item, then moves the mouse to select the Paraglider item, then moves the mouse to select the Ram-Hornize item, and finally moves the mouse to explore the wardrobe options.||", "The user moves the mouse slightly up and to the left, then clicks the left mouse button twice.||"

Zoom in on Annotations

Generated Data - NMS

```
MouseRight=4+KeyW,
KeyW,
KeyW,
MouseRight=4+MouseUp=5+KeyW
```

The player turns slightly to the right and moves forward three steps.

The player then turns slightly to the right again and moves forward one more step ...

while releasing the mouse button.

Generated Data - Valheim

```
MouseDown=3,
MouseDown=3,
MouseDown=2,
KeyW,
KeyW,
KeyW,
KeyW,
KeyW,
KeyW
```

The player moves the mouse slightly down and to the right.

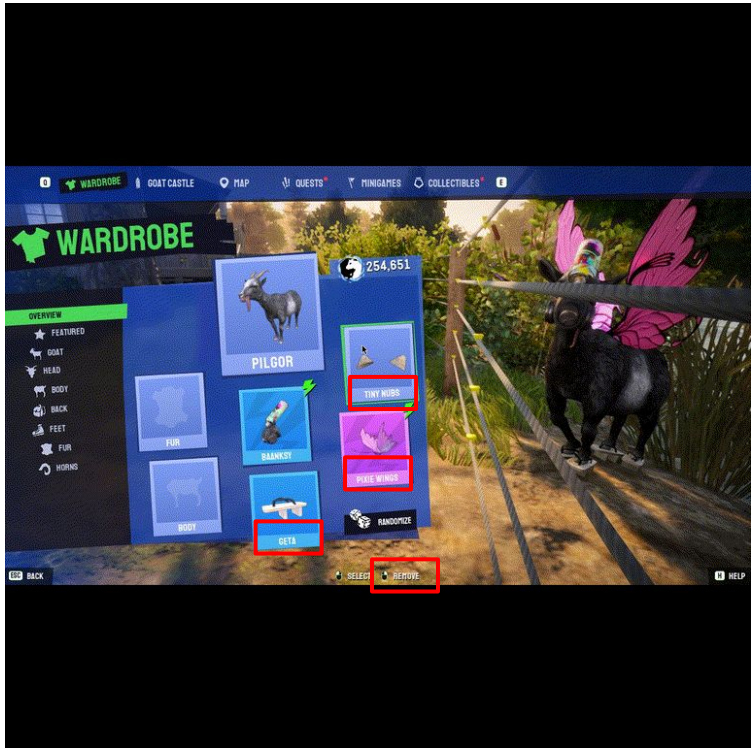
The player then presses the W key six times to move forward.

Generated Data - Valheim Menu use

{
Tab,
Tab,
Tab,

The player presses the Tab key three times to navigate the crafting menu.

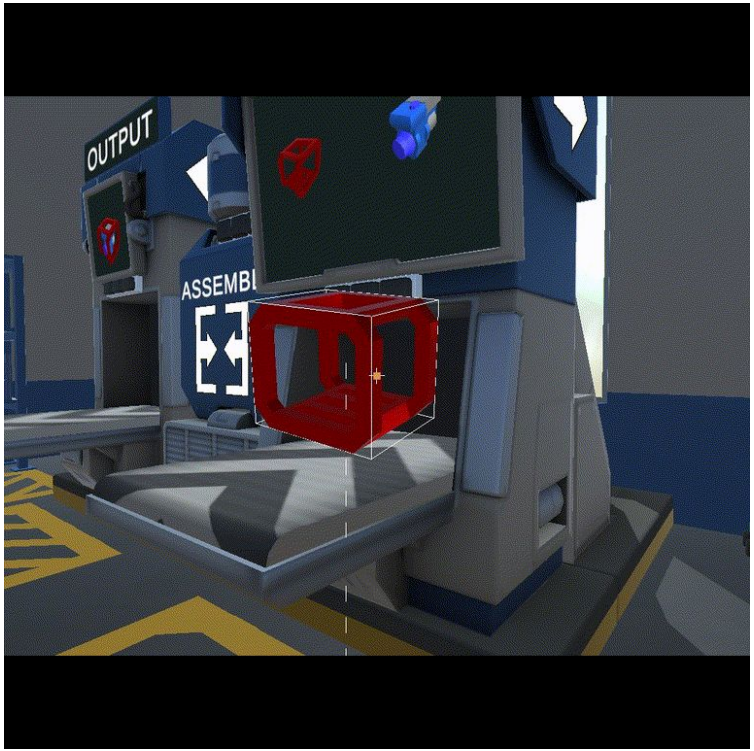
Generated Data - GS3




```
MouseLeft=2+MouseUp=1,
MouseLeft=13+MouseDown=2,
MouseLeft=3+MouseUp=2,
MouseLeft=180+MouseDown=4
```

The user moves the mouse to select the **Pixie Wings**, then moves the mouse to select the **Geta**, and finally moves the mouse to remove the **Tiny Nubs**.

Generated Data - GS3



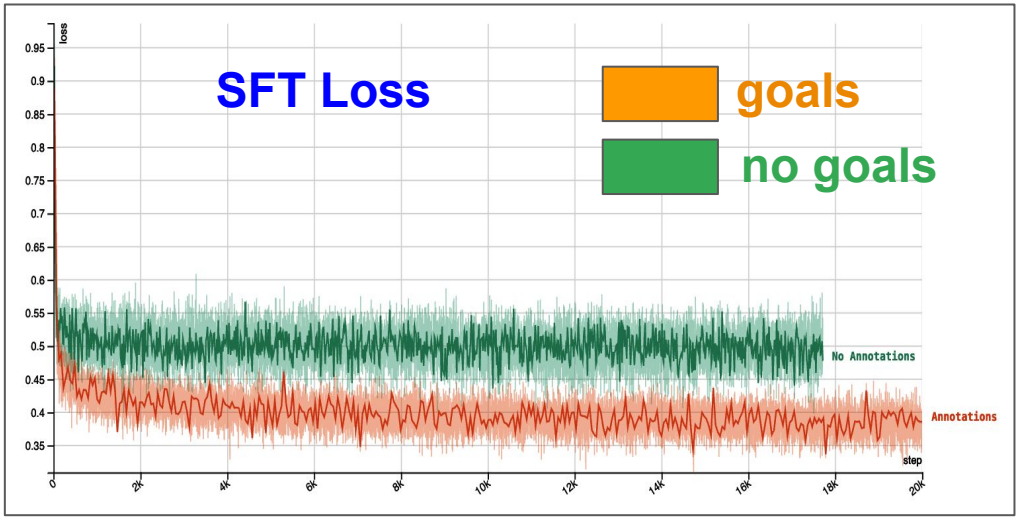

```
ClickLeft,
ClickLeft,
MouseLeft=1+MouseUp=1,
MouseLeft=12+MouseDown=9
```

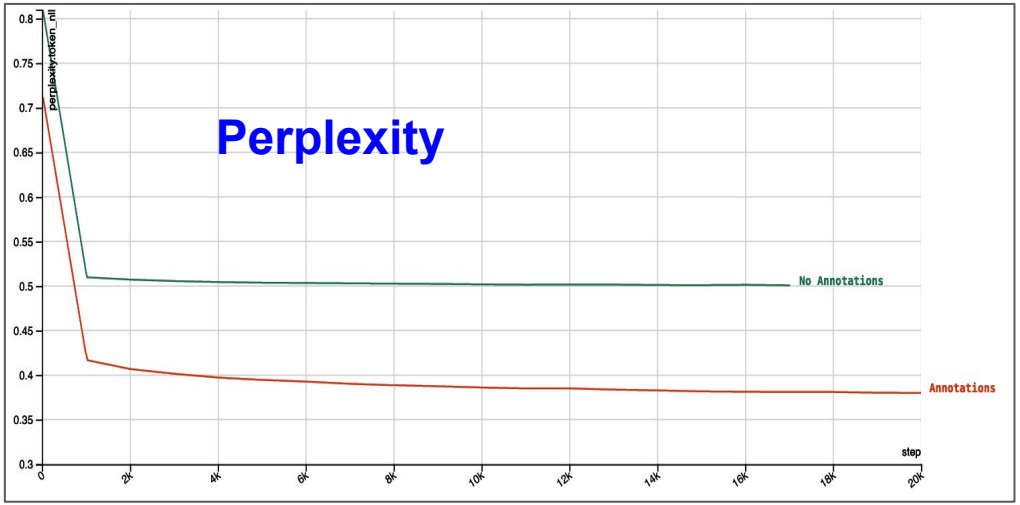
The player clicks the red box to pick it up, then releases it after moving the mouse slightly to the left.

Fine Tuning & Inference

Fine Tune a Model

- ~1.2M Spans based on 80K Canonical datasets
- Annotation / no-annotation datasets composed of the same spans
- S V2 Model
- Used simple (but likely non-optimal) key frame criteria





In Distribution Inference - Valheim (@20 steps)

Your next instruction is:
Go close to the wood door.

Output
7.0s

Assistant

action-goals

The player moves forward four steps.||KeyW,KeyW,KeyW,KeyW

Output
8.7s

Assistant

no action-goals

MouseLeft=62+MouseUp=15,MouseLeft=29+MouseUp=8,

MouseLeft=22+MouseUp=13,MouseLeft=16+MouseUp=11

The instruction and the goal here seem mostly consistent. The baseline model actions don't appear to align too well with the instruction.

In Distribution Inference - GS3 (@20K steps)

Your next instruction is:
Run up the path.

Output
10.4s

Assistant

The user presses the W key while holding down the left shift key four times to move forward.||KeyW+ShiftLeft,KeyW+ShiftLeft,KeyW+ShiftLeft,KeyW+ShiftLeft

Ground truth data:

```
KeyW+ShiftLeft,
KeyW+ShiftLeft,
KeyW+ShiftLeft,
KeyW+ShiftLeft
```

A simple case where goals and actions align and do so with ground truth.

In Distribution Inference - GS3 (@20K steps)

Your next instruction is:
Move towards the right.

Output
8.3s

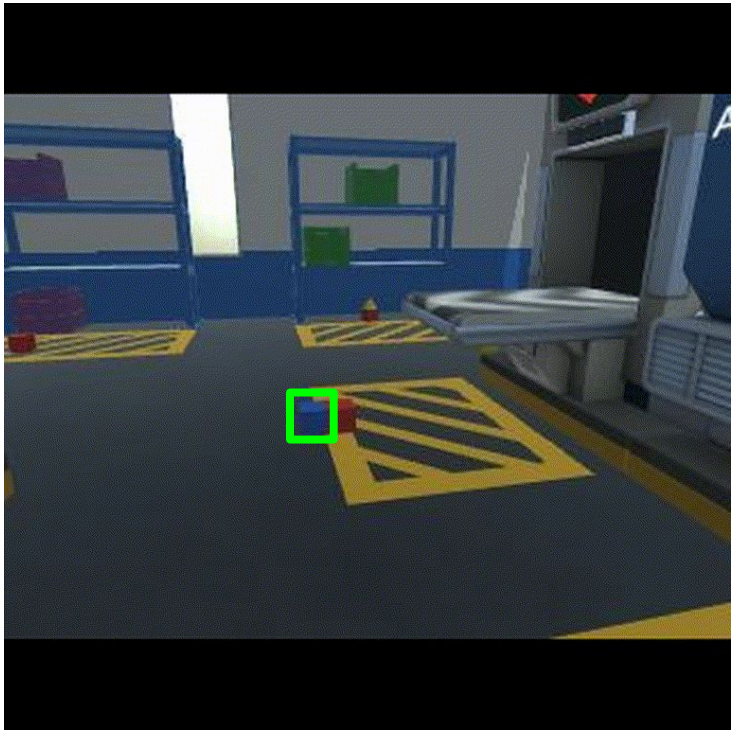
Assistant

The player moves the mouse slightly to the right while pressing the D key to move to the right, repeating this action three times.||MouseLeft=1+KeyD,MouseLeft=4+KeyD,MouseLeft=5+KeyD,MouseLeft=17+KeyD

Interesting case, the model goal seems to align well to the instruction however the action doesn't as it appears to be veering left.

It's possible that these are the ground truth actions but the model is attempting to correct for this in the goals. Will follow up.

OOD[instruction] Inference - Factory (@20K steps)



Your next instruction is:
Pick up the blue cube.

Output 8.4s

Assistant

action-goals

The player clicks the left mouse button to pick up the box, then moves the mouse slightly to the left and up to lift the box.||ClickLeft,MouseLeft=7+MouseUp=10,MouseLeft=11+MouseUp=13,MouseLeft=21+MouseUp=15

Output 8.8s

Assistant

no action-goals

MouseRight=23+MouseDown=43,MouseRight=1+MouseDown=4,MouseDown=1,MouseDown=1

The goal model attempts to pick up an object on the left, semi consistent to the proposed instruction.

OOD[in-domain] Inference - Valheim (@20K steps)

Your next instruction is:
Run to the green inner tube.

Output 28.5s — Assistant — **action-goals**

The user presses the W key four times to move the goat forward.||KeyW,KeyW,KeyW,KeyW

Output 51.4s — Assistant — **no action-goals**

MouseLeft=29+MouseDown=3,MouseLeft=32,MouseLeft=13+MouseUp=1,MouseLeft=19+MouseUp=6

Neither are ideal but the goals model seems to do a it better toward the instruction here.

OOD[domain] Inference - Super Mario (@20K steps)

The goal model expresses the intent to move forward a significant amount. The baseline moves the mouse, presumably to look around.

The former should likely get us further along to the overall goal.

Your next instruction is:
go to the yellow box

Output 7.6s

Assistant

action-goals

The player presses the W key repeatedly to move forward.||KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW,KeyW

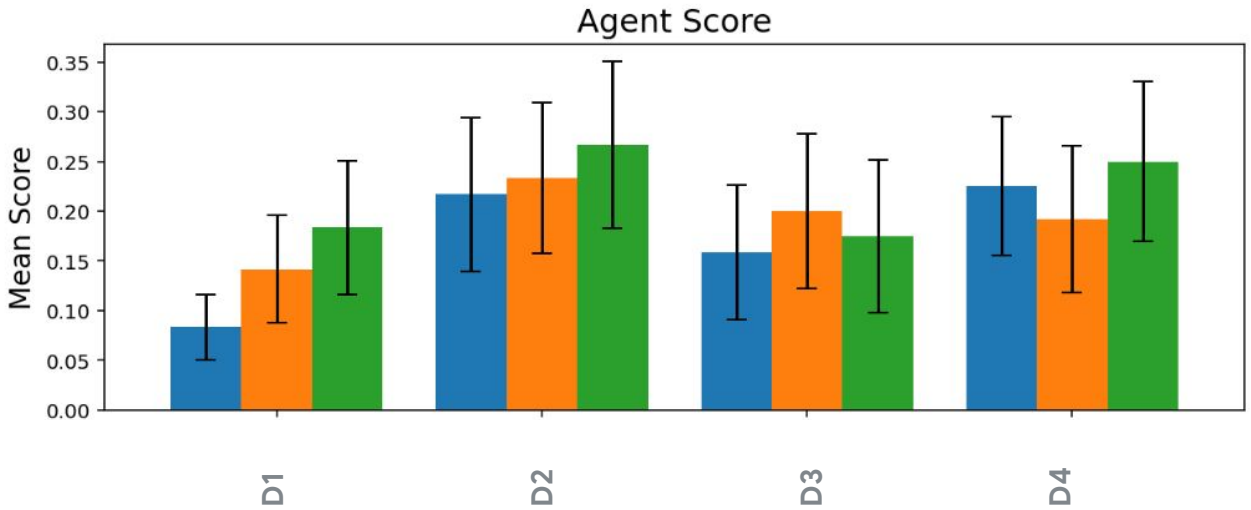
Output 7.0s

Assistant

no action-goals

MouseLeft=36+MouseDown=3+KeyS,MouseLeft=20+KeyS,MouseLeft=76+MouseDown=2+KeyS,MouseLeft=75+KeyS

Scores - S Model, 6 Runs per Domain



17.1	Baseline @20K
19.2	Goals @20K
21.9	Goals @40K

Next Steps

This is a first draft of annotations so we can improve these:

- Lots of repetition in the annotations. The data contains a number of small segments.
- Use the annotation model to also pick the key frames. The aim is to partition the span into natural sub-sequences of actions.
- More data from a better source (canonicals are known to be sub-optimal).
- Integrate with action groupings
- Evaluate on longer horizon tasks. Make use of goldmine style interface?