Multi-agent hide and seek

Тарасов Денис Алексеевич

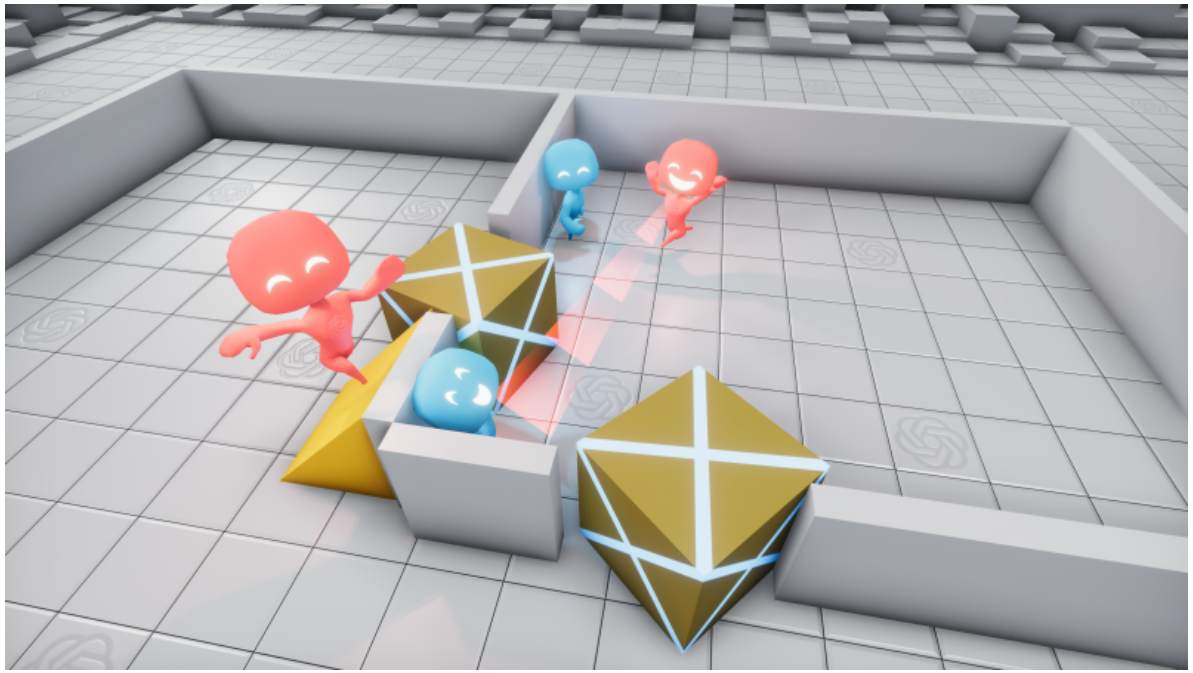
Ментор: Екатерина Смаль (JetBrains Research)

Санкт-Петербургская школа физико-математических и компьютерных наук

НИУ ВШЭ – Санкт-Петербург

Описание проекта

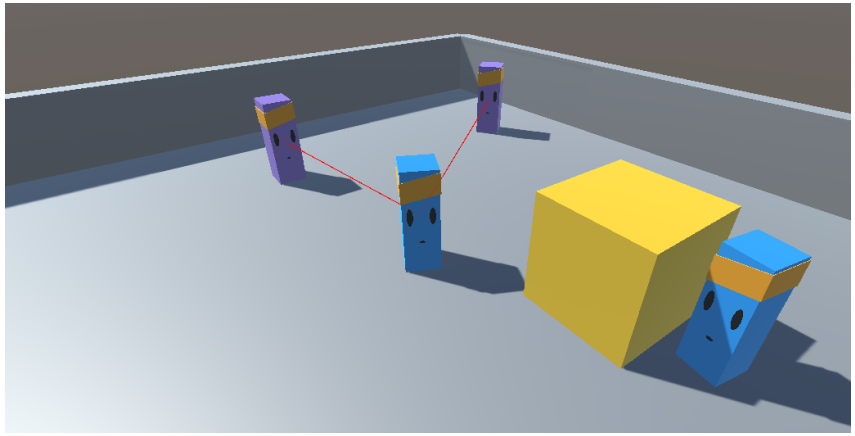
В конце 2019 года компания OpenAI опубликовала статью[1], где агенты при игре в прятки научились использовать вспомогательные предметы (коробки, рампы). Задача практики – повторить результат статьи с использованием бесплатного движка Unity, вместо комерческого MuJoCo[3].



Простая среда для игры из работы OpenAI

Игра

- Две команды агентов: прячущиеся (синие) и ищущие (фиолетовые)
- Агенты на основе своих наблюдений совершают действия
- Если ищущие видят хотя бы одного прячущегося, то все ищущие получают положительную награду, все прячущиеся отрицательную, иначе наоборот
- Цель каждого агента – максимизировать свою награду
- Поведение агентов меняется по средствам обучения с подкреплением



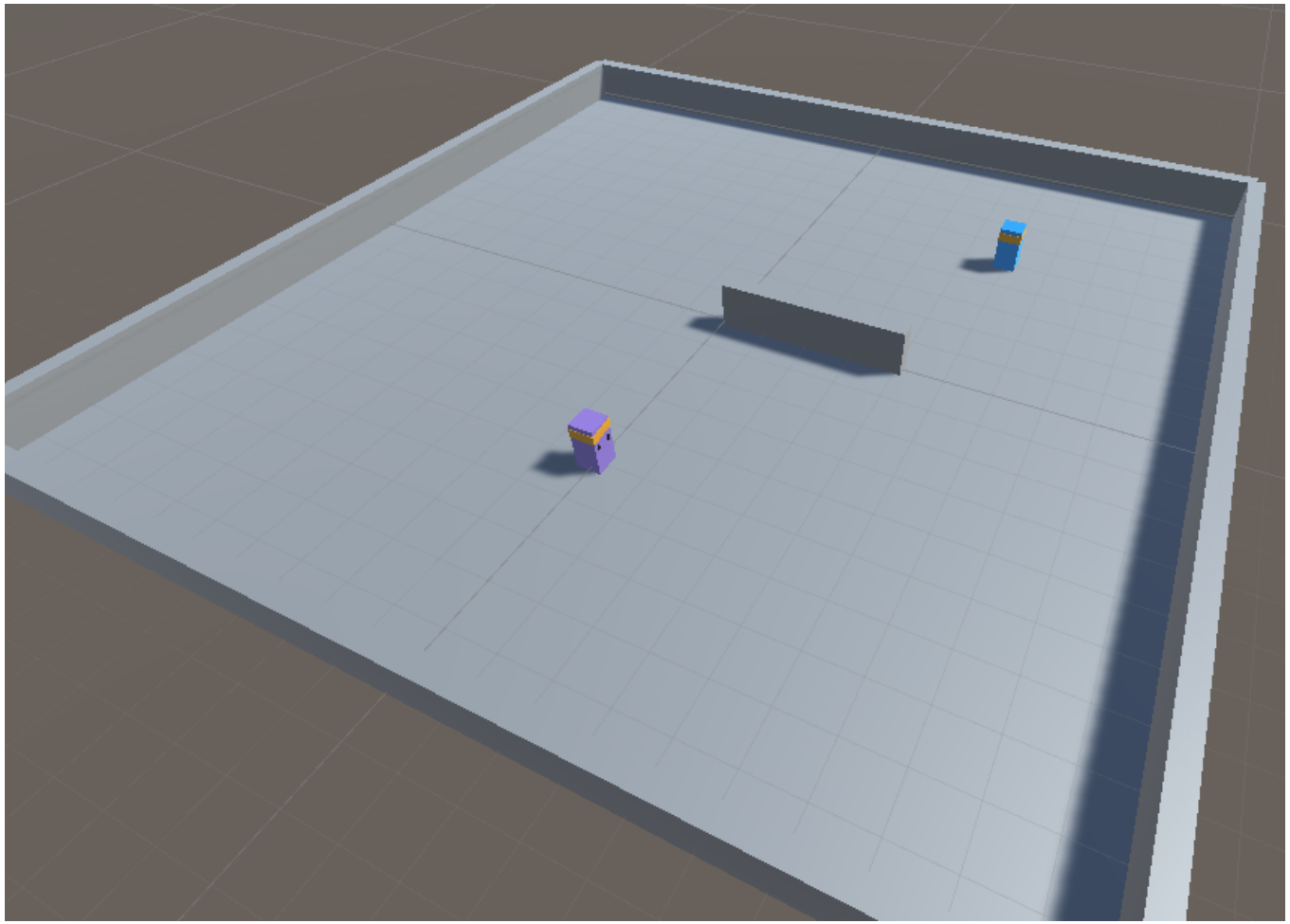
Используемые технологии

- C#
- Unity
- Unity ML-Agents Toolkit [2]

Этапы

- Создать простую среду и попробовать обучить агентов в ней
- Итеративно усложнять среду и агентов

Среда 1



Среда с одним агентом в каждой команде и одной стеной

Среда 1

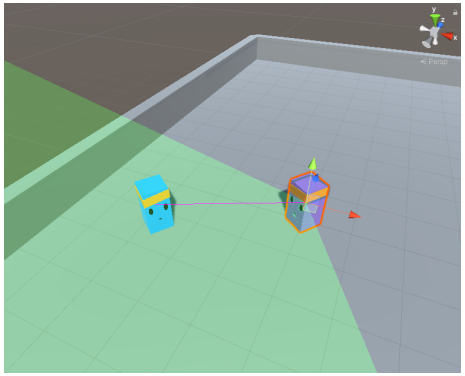
Агенты

Наблюдения:

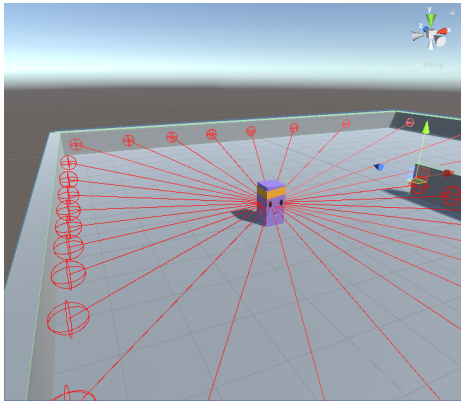
- $\vec{x}$, $\vec{v}$ себя
- $\vec{x}$, $\vec{v}$ оппонента, если он попал в ”поле зрения”
- ”Лидар”

Действия:

- Применить силу к себе
- Повернуться вокруг оси y



Поле зрения



Лидар

Среда 1

Результаты обучения

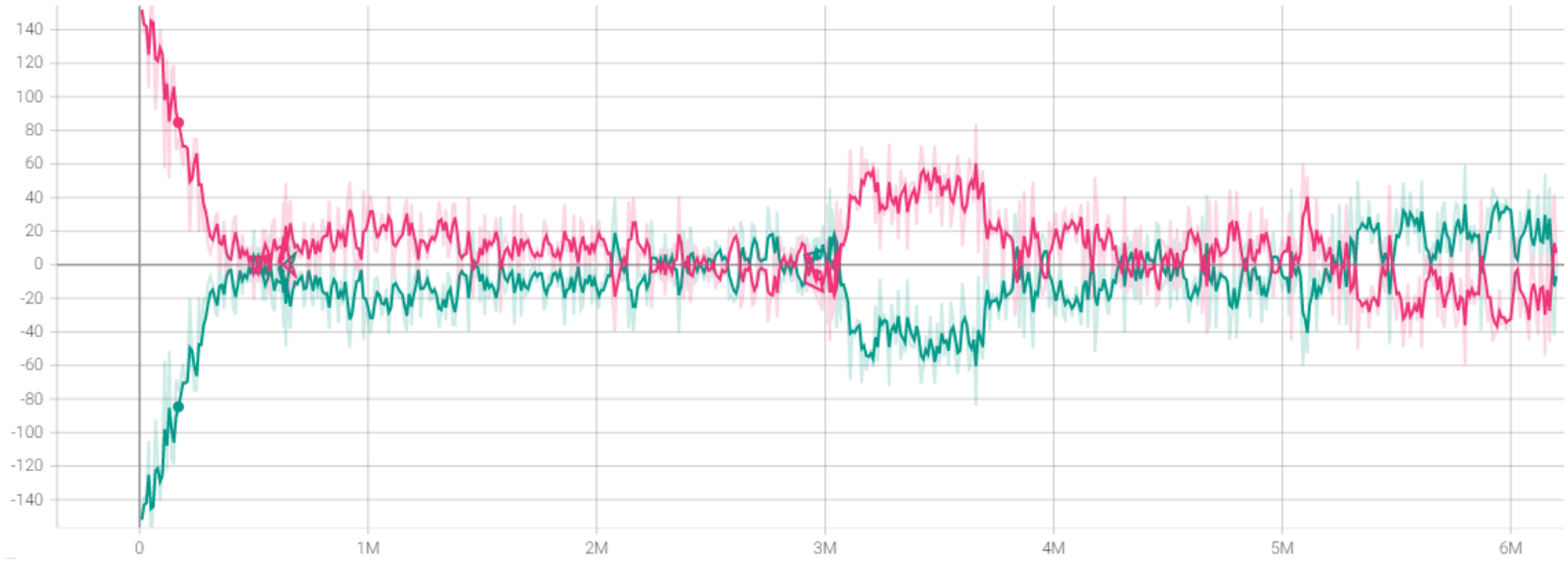
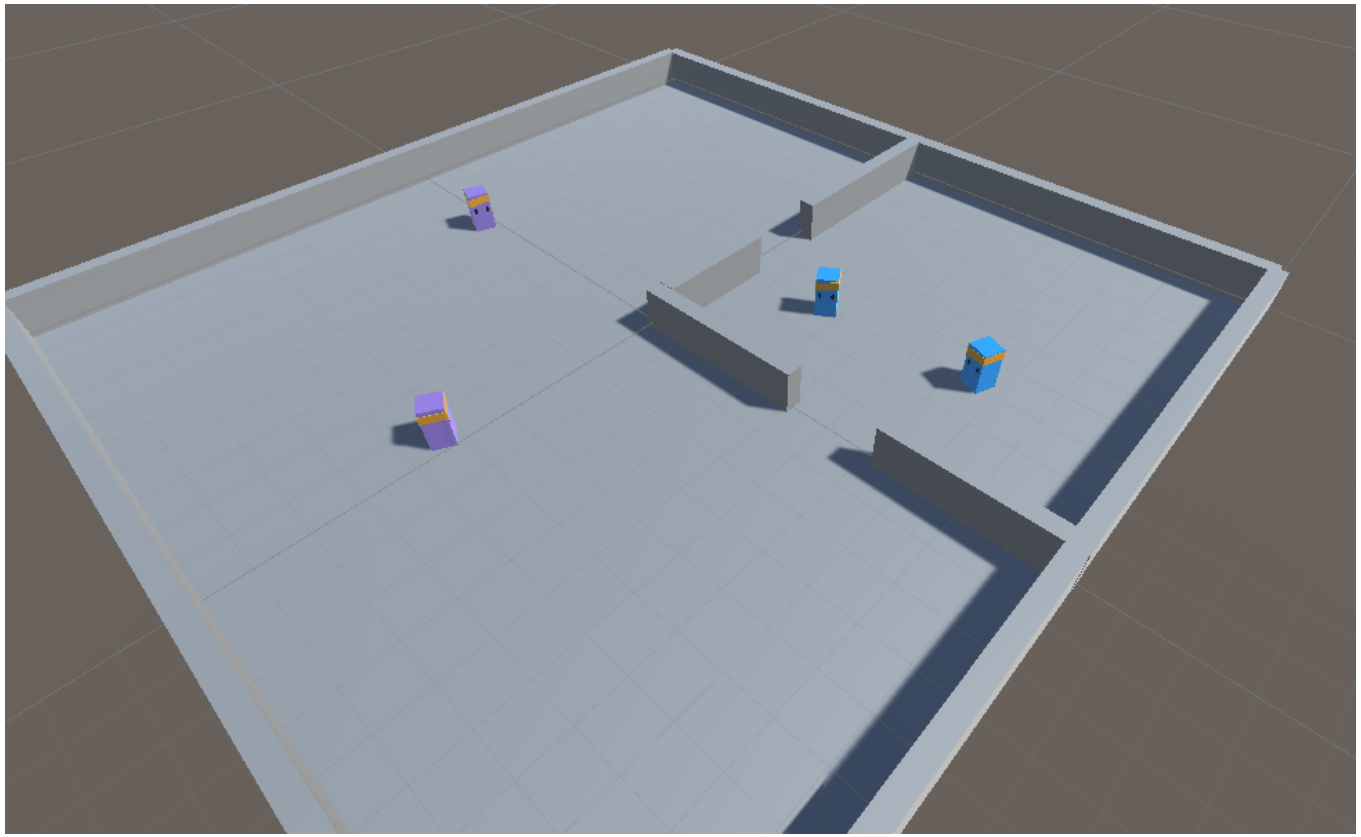


График обучения. По горизонтали – количество шагов симуляции, по вертикали – средняя награда, зеленый – ищущий, красный – прячущийся

https://youtu.be/G6lHAKiyQIE

Среда 2



Среда с двумя агентами в каждой команде и фиксированной комнатой

Среда 2

Агенты

Наблюдения:

- $\vec{x}$, $\vec{v}$ себя
- $\vec{x}$, $\vec{v}$, команда других агентов в ”поле зрения”
- ”Лидар”

Действия:

- Применить силу к себе
- Повернуться вокруг оси y

Среда 2

Результаты обучения

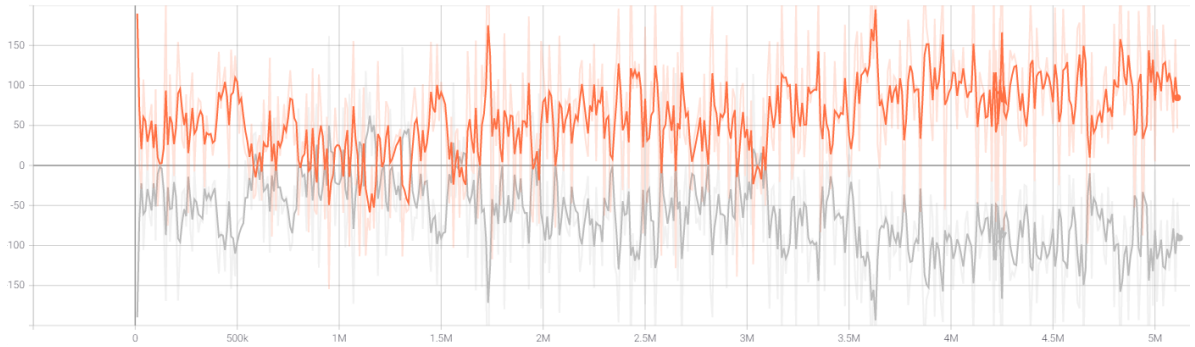


График обучения со стадией подготовки. Серый – ищущие, оранжевый – прячущиеся

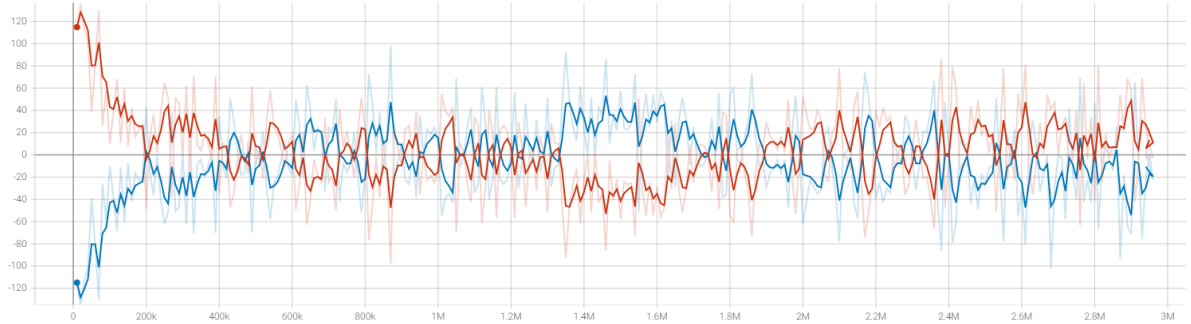
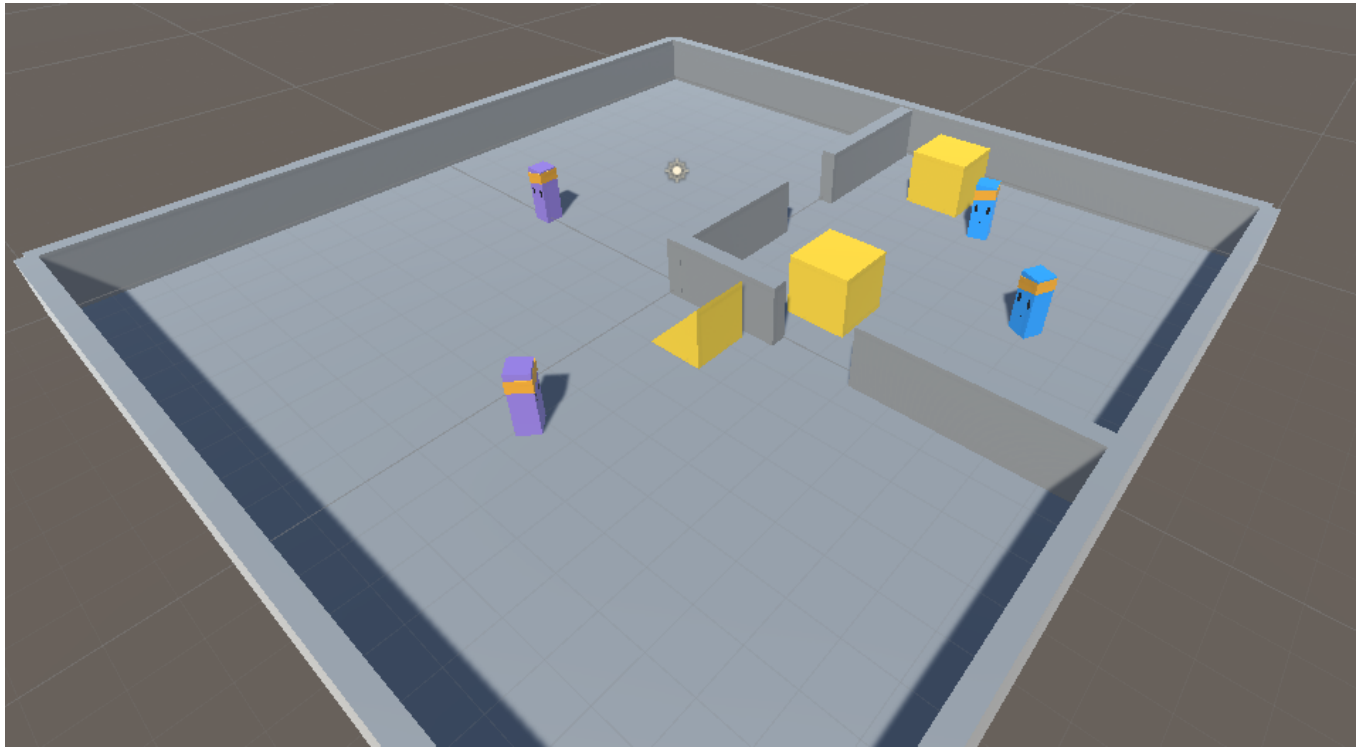


График обучения без стадии подготовки. Синий – ищущие, красный – прячущиеся

Среда 3



Среда с двумя агентами в каждой команде, случайным расположением отверстий в стенах комнаты и интерактивными объектами

Среда 3

Агенты

Наблюдения:

- $\vec{x}$, $\vec{v}$ себя
- $\vec{x}$, $\vec{v}$, тип других движущихся объектов в ”поле зрения”
- ”Лидар”
- Оставшееся время до конца подготовки

Действия:

- Применить силу к себе
- Повернуться вокруг оси y
- Взять/отпустить объект
- Заблокировать/разблокировать объект

Среда 3

Результаты и демонстрация функционирования среды

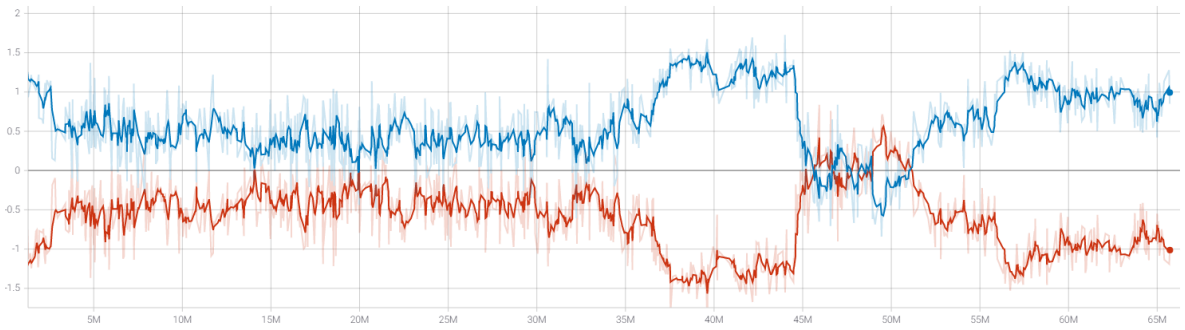


График обучения. По горизонтали – количество шагов симуляции, по вертикали – средняя награда, красный – ищущий, синий – прячущийся

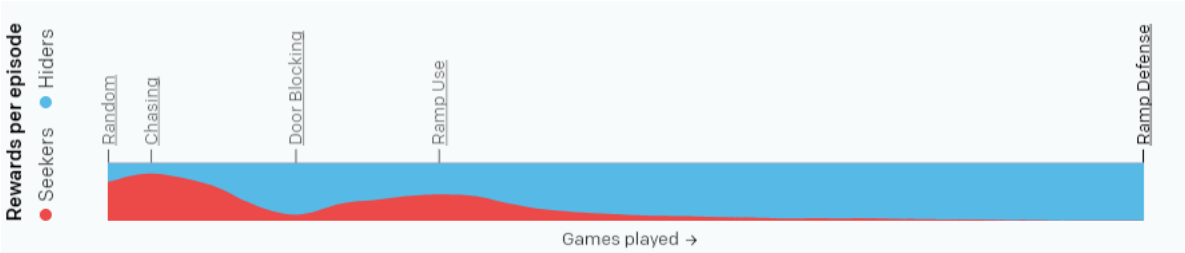


График обучения из блога OpenAI

https://youtu.be/Bk1vIVzacZs

Дальнейшая работа

- Обучение агентов в 3-й версии среды
- Воспроизведение ”полной” среды из статьи
- Воспроизведение тестов на ”интеллектуальность” обученных агентов

Репозиторий

https://github.com/DT6A/MAHS

Ссылки

- Bowen Baker et al. “Emergent Tool Use From Multi-Agent Autocurricula.” In: *CoRR* abs/1909.07528 (2019). arXiv: 1909.07528. URL: http://arxiv.org/abs/1909.07528.
- Arthur Juliani et al. “Unity: A General Platform for Intelligent Agents.” In: *CoRR* abs/1809.02627 (2018). arXiv: 1809.02627. URL: http://arxiv.org/abs/1809.02627.
- Emanuel Todorov, Tom Erez, and Yuval Tassa. “MuJoCo: A physics engine for model-based control.” In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012, pp. 5026–5033. DOI: 10.1109/IROS.2012.6386109.