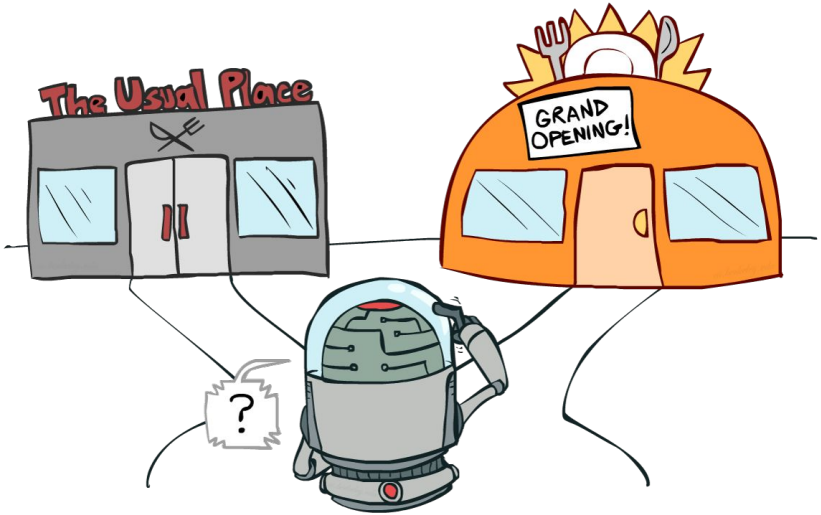
Improving Reinforcement Learning through Natural Language Processing

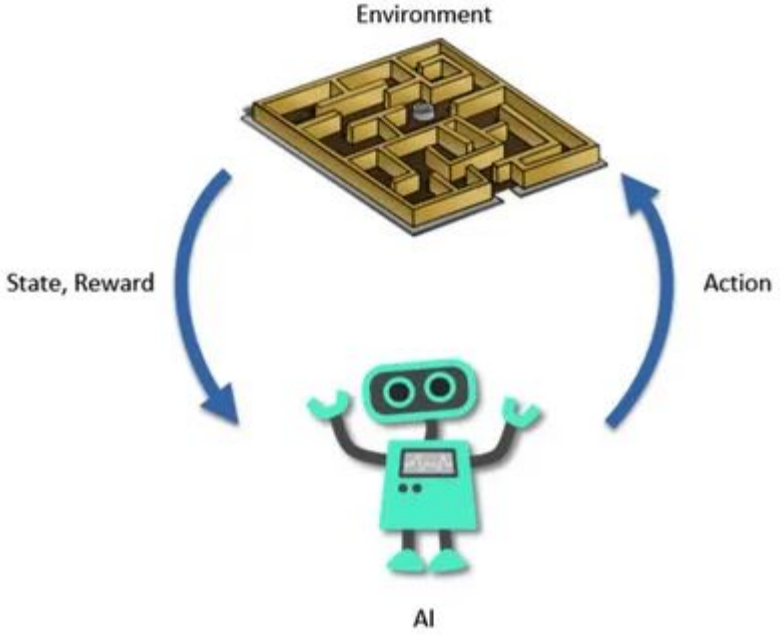
Выполнил:
Разин Арслан Дмитриевич, БПМИ202

1. **Введение**
2. VLN
3. LEARN
4. LangLfP
5. LLM for reward
6. Выводы
7. Источники

Проблемы обучения с подкреплением



Exploration vs exploitation problem



Reward problem

Как можно решить эти проблемы?

Video and text inputs

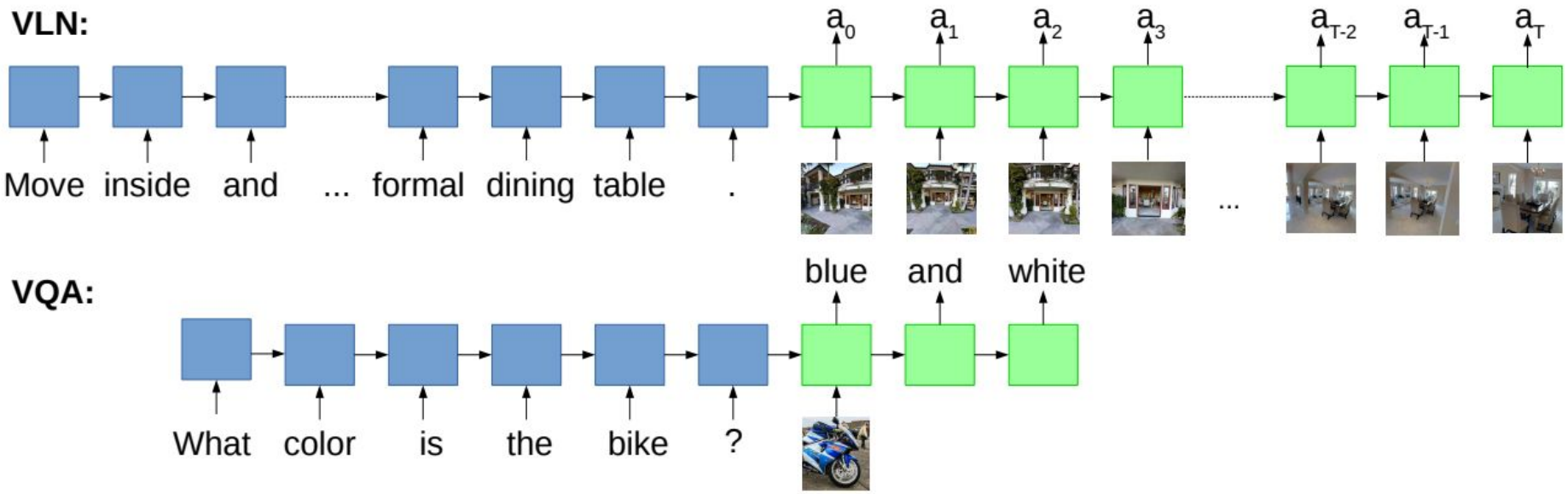
the **bottle** is in the living room

get the **bottle**

the **plates** are in the

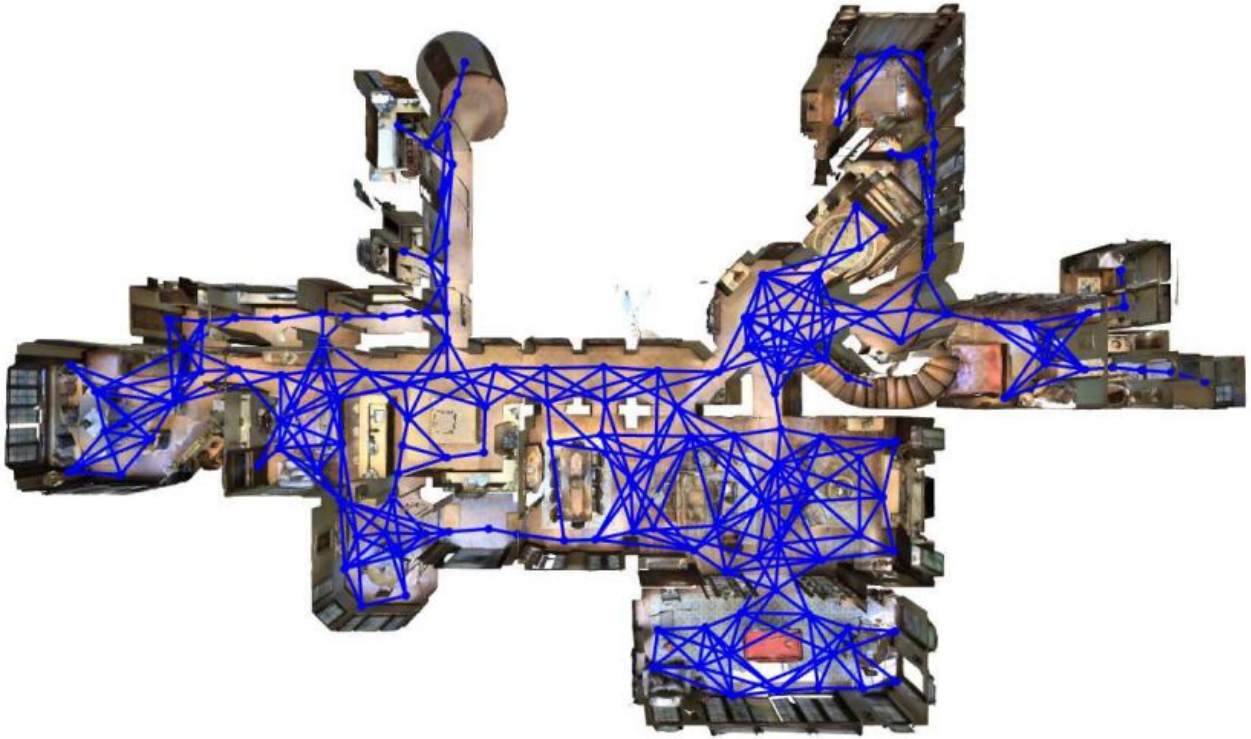
1. Введение
2. **VLN**
3. LEARN
4. LangLfP
5. LLM for reward
6. Выводы
7. Источники

Vision-and-Language Navigation



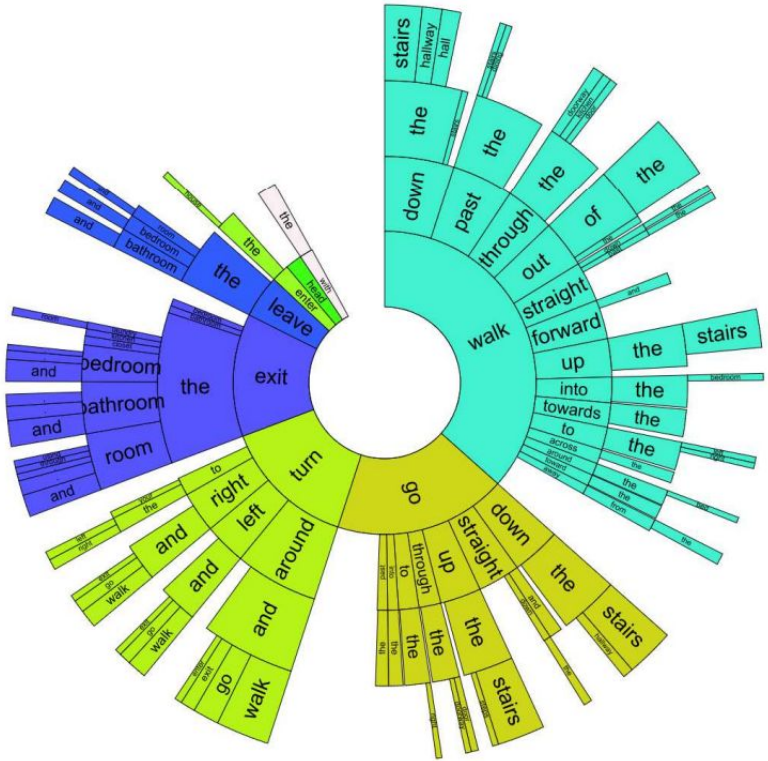
Процесс обработки инструкции роботом

Vision-and-Language Navigation



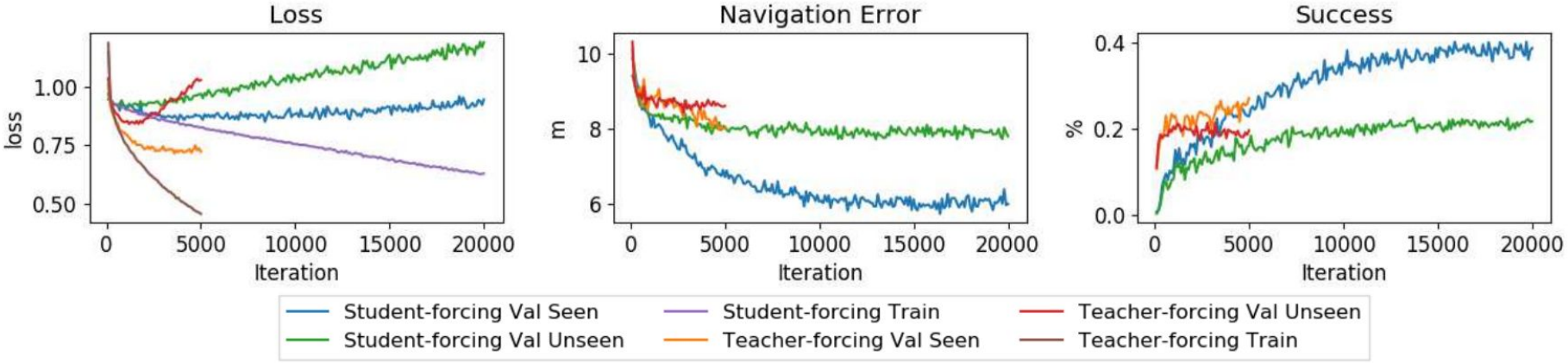
Пример данных из Room-2-Room датасета

Vision-and-Language Navigation



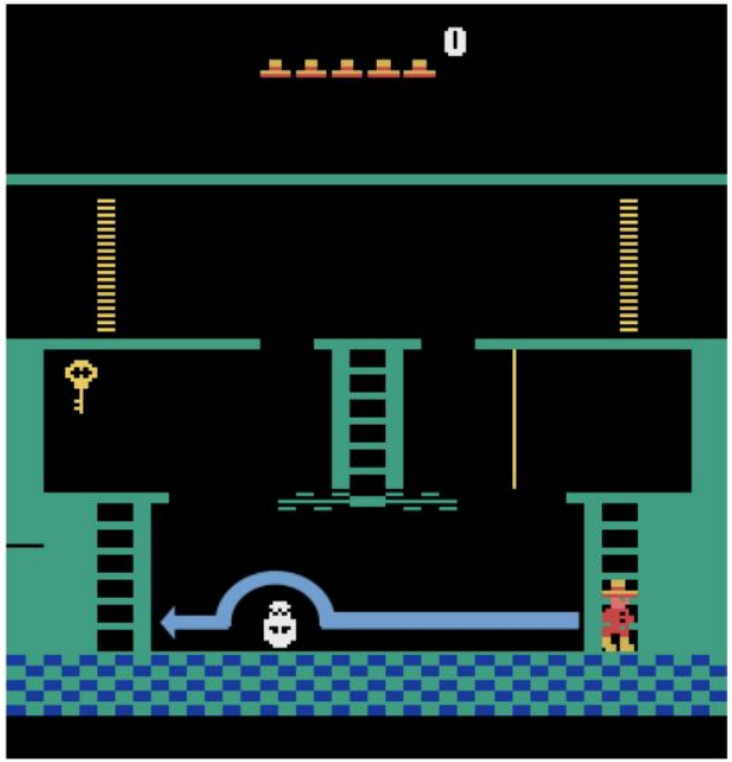
Представление инструкции по первым 4 словам

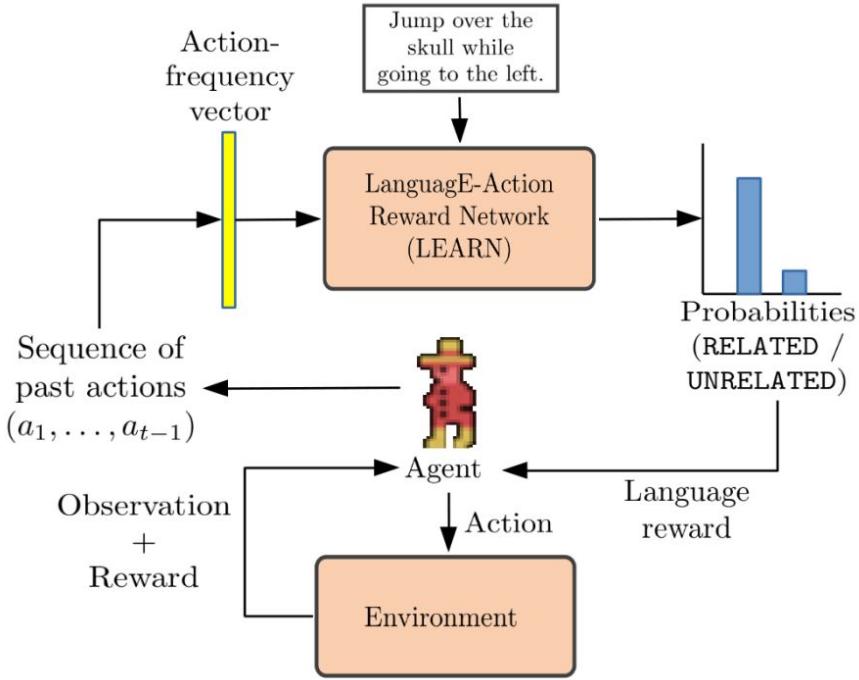
Vision-and-Language Navigation



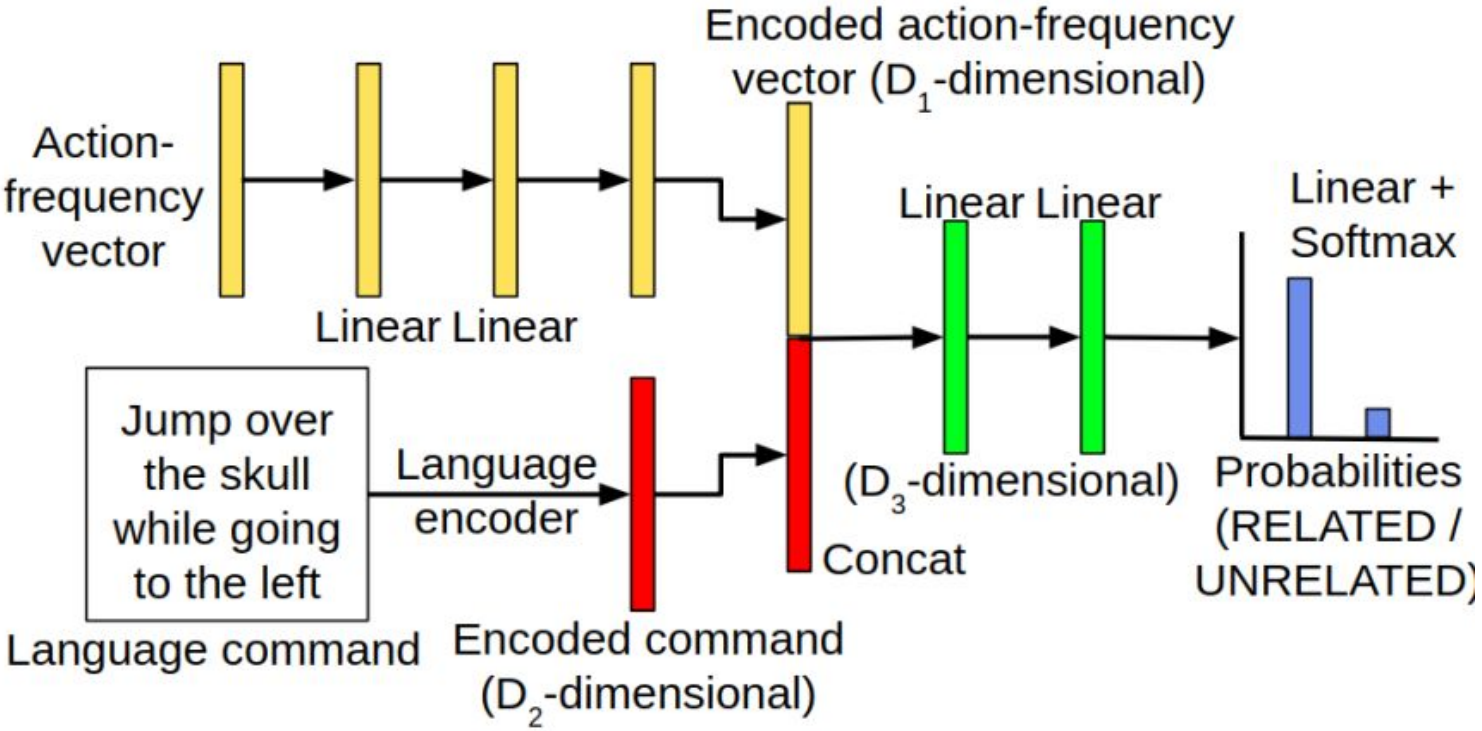
1. Введение
2. VLN
3. **LEARN**
4. LangLfP
5. LLM for reward
6. Выводы
7. Источники

LanguagE-Action Reward Network

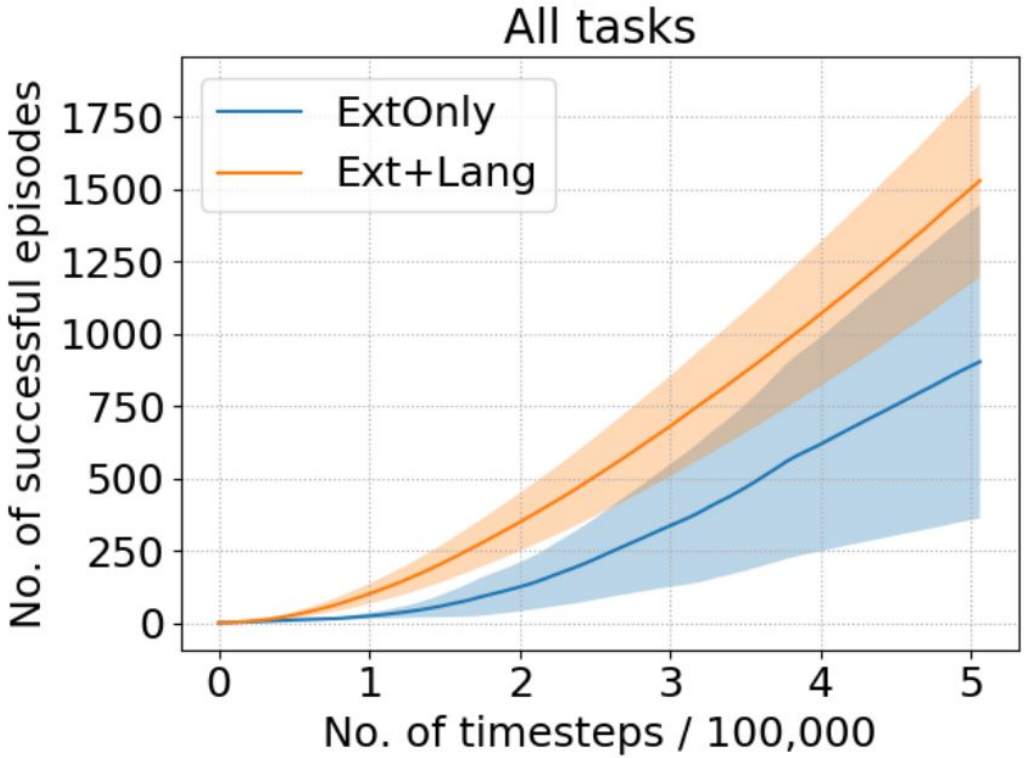




LanguagE-Action Reward Network

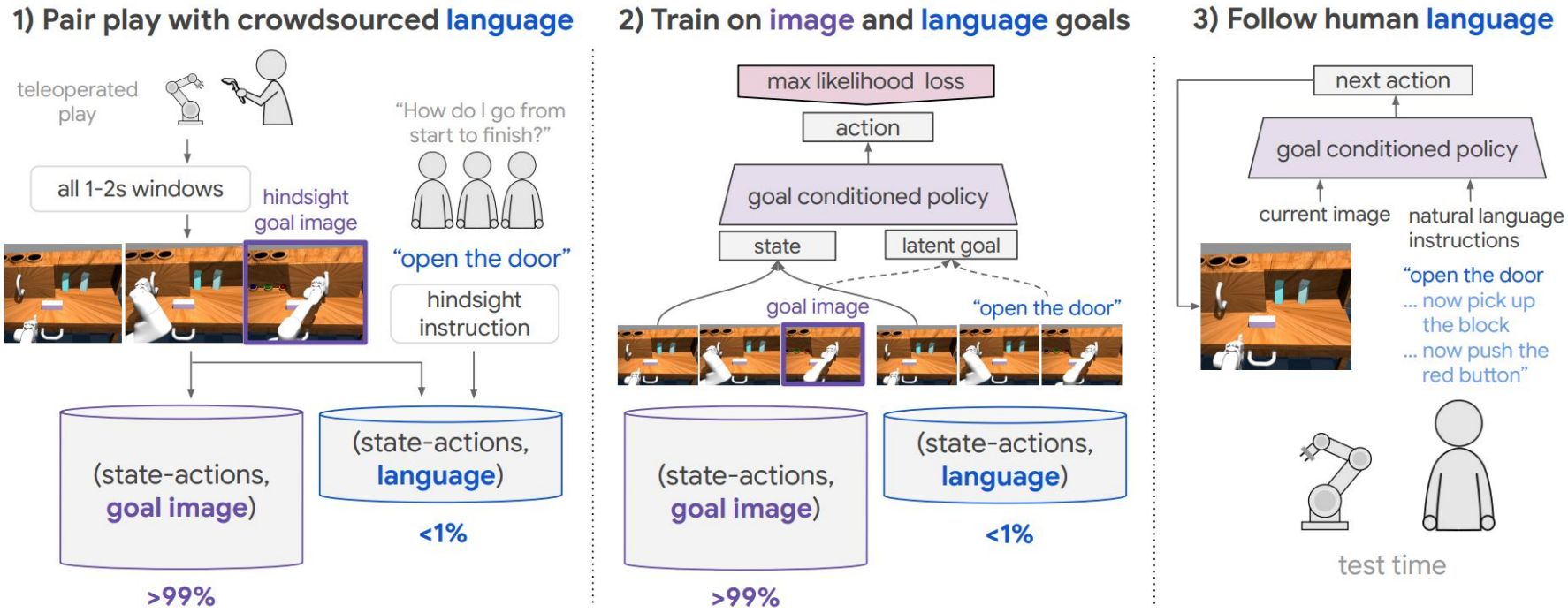


LanguagE-Action Reward Network

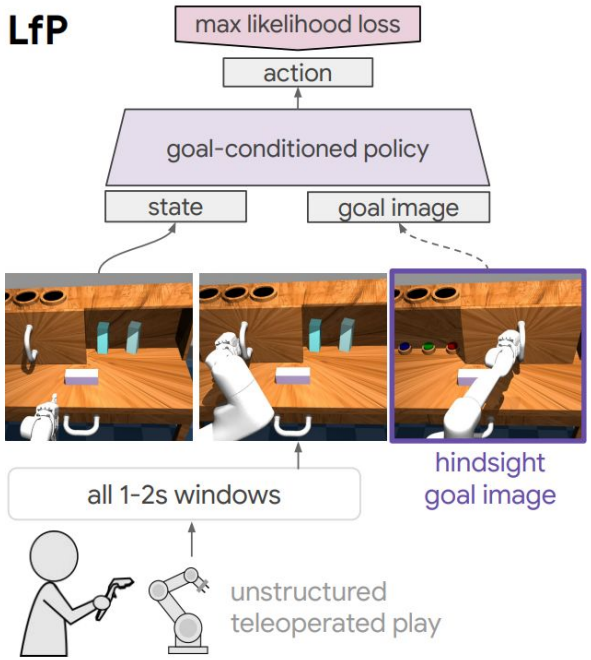


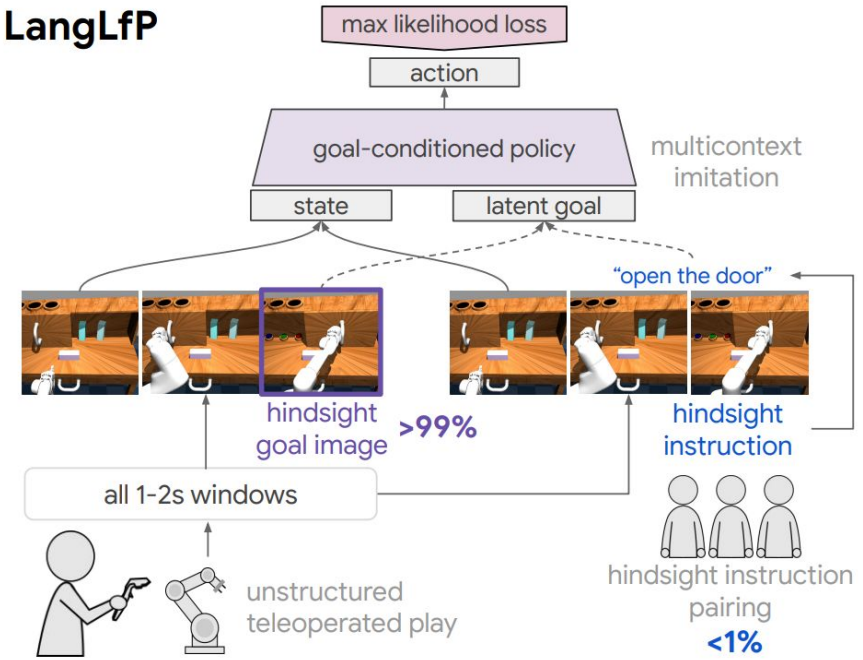
1. Введение
2. VLN
3. LEARN
4. **LangLfP**
5. LLM for reward
6. Выводы
7. Источники

Language Learning from Play

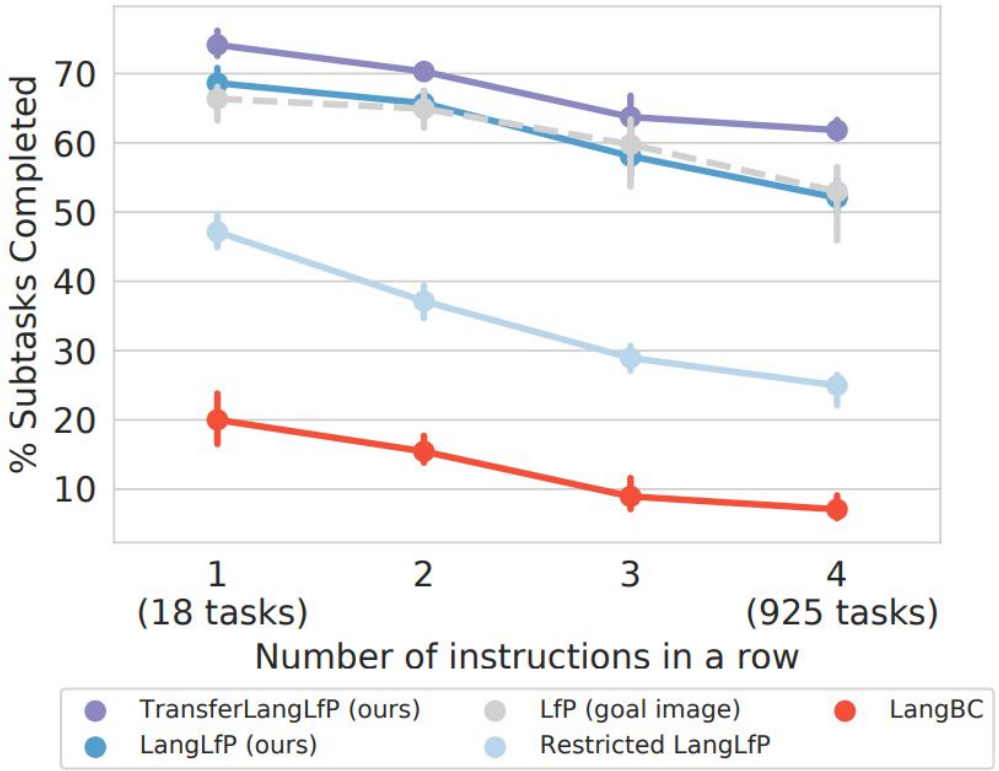


Language Learning from Play



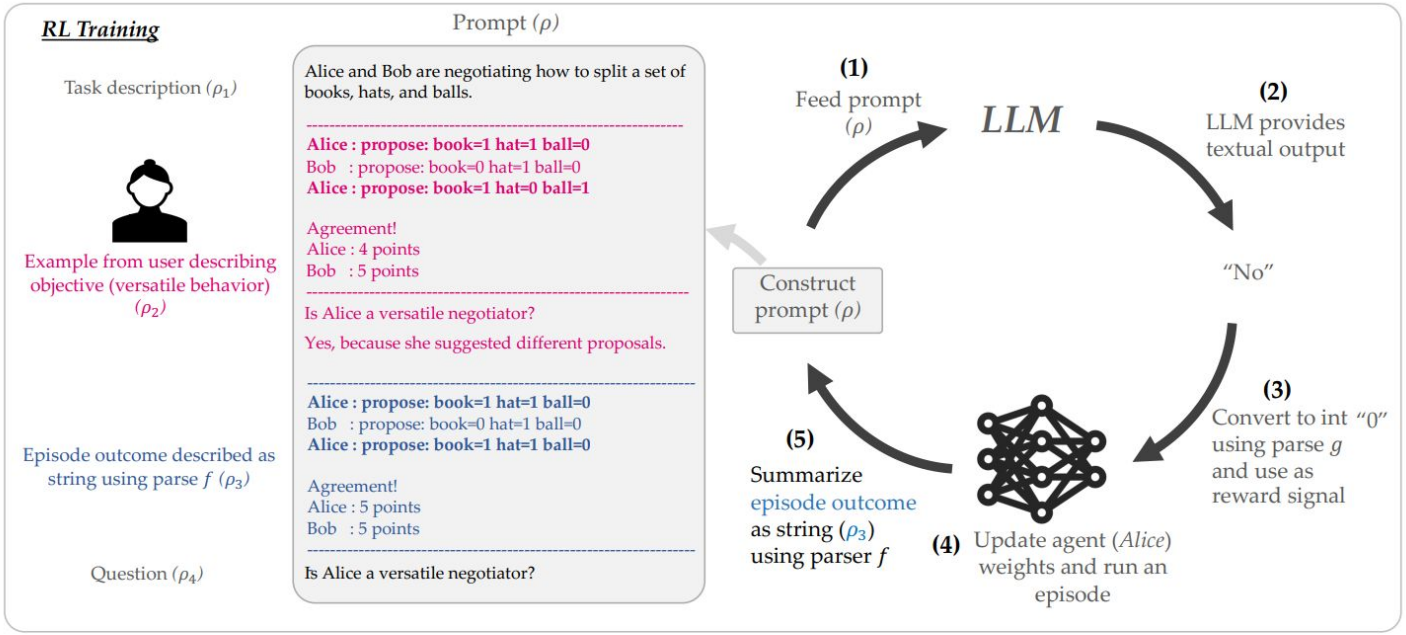


Language Learning from Play



1. Введение
2. VLN
3. LEARN
4. LangLfP
5. **LLM for reward**
6. Выводы
7. Источники

LLM for reward



LLM for reward

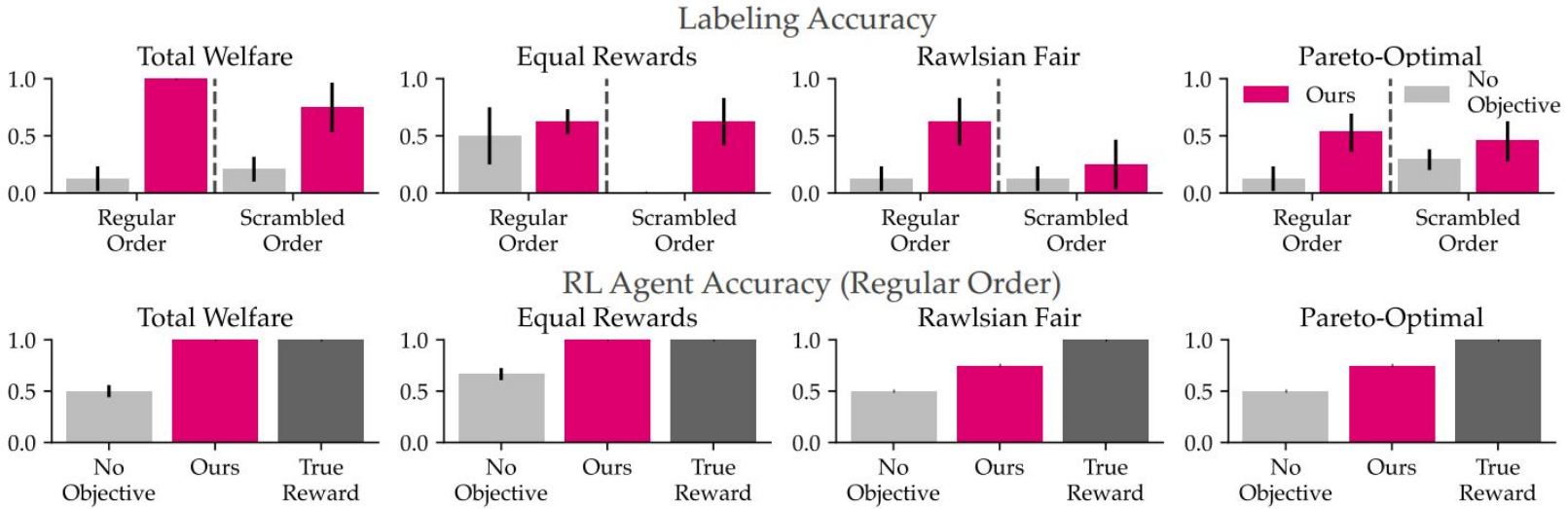


Figure 3: **Matrix Games, Zero-shot**. (Top) Accuracy of reward signals provided by LLM and a *No Objective* baseline during RL training. We report results for both regular and scrambled versions of matrix games. (Bottom) Accuracy of RL agents after training.

LLM for reward

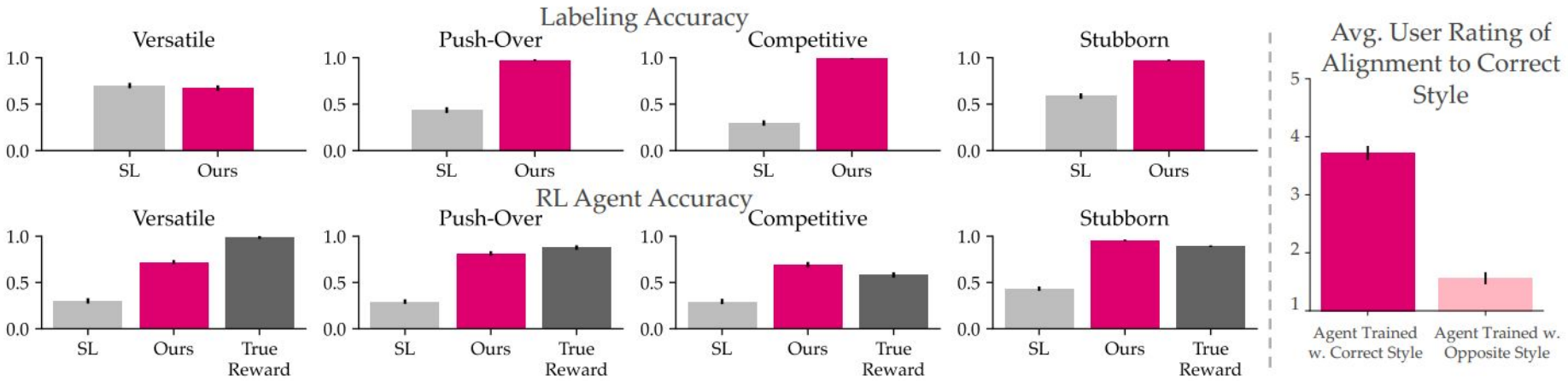


Figure 4: **DEALORNODEAL, Few-shot**. (Top) Accuracy of reward signals provided by LLM and SL during RL training. (Bottom) Accuracy of RL agents after training. (Right) Pilot study results. Agents trained with the user's preferred style were rated as significantly more aligned than an agent trained with the opposite style $p < 0.001$.

1. Введение
2. VLN
3. LEARN
4. LangLfP
5. LLM for reward
6. **Выводы**
7. Источники

Выводы

Использование NLP в рамках RL открывает новые перспективы для повышения эффективности и гибкости обучения агентов:

- NLP может быть использовано для дизайна системы наград, где комплексные задачи описываются на естественном языке, позволяя агенту интерпретировать цели и промежуточные задачи в более широком контексте.
- Кроме того, естественный язык может служить средством для предоставления подсказок и инструкций, что упрощает процесс обучения агента и помогает ему быстрее адаптироваться к новым условиям и задачам.

Такой подход к обучению агентов в первую очередь важен для робототехники, но также помогает в более RL классических задачах.

1. Введение
2. VLN
3. LEARN
4. LangLfP
5. LLM for reward
6. Выводы
7. **Источники**

Источники

Основная статья: https://arxiv.org/pdf/2308.01399.pdf

Статьи, из которых взяты примеры:

- VLN
- LEARN
- LangLfP
- LLM for reward