

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**Федеральное государственное автономное образовательное**  
**учреждение высшего образования**

**Национальный исследовательский университет**  
**«Высшая школа экономики»**

**Факультет гуманитарных наук**  
**Образовательная программа**  
**«Фундаментальная и компьютерная лингвистика»**

**КУРСОВАЯ РАБОТА**

На тему «Общение ИИ агентов»

*Тема на английском «Communication of AI Agents»*

Студент 3 курса группы №211

Хайрутдинов Тагир Рамилевич  
(Ф.И.О.)

Научный руководитель

Клышинский Эдуард Станиславович  
(Ф.И.О.)

Доцент, Школа Лингвистики

(должность, звание)

Научный консультант

Сериков Олег Алексеевич  
(Ф.И.О.)

Приглашенный преподаватель

(должность, звание)

Москва, 2024 г.

## Оглавление

<b>1. Введение.....</b>	<b>2</b>
1.1. Обзор темы.....	2
1.2. Постановка задачи.....	3
1.3. Обзор литературы.....	4
<b>2. Методы и материалы.....</b>	<b>5</b>
2.1. Что такое многорукие бандиты.....	5
2.2. Элементы RL в задачах многоруких бандитов.....	6
2.3. Стратегии для многоруких бандитов.....	7
2.4. Методы оценки эффективности взаимодействия агентов.....	8
2.5. Описание ролей агентов и их задач.....	9
2.6. Описание режимов игры.....	10
2.7. Написание кода для проведения экспериментов.....	12
2.8. Дизайн экспериментов.....	13
2.9. Методология оценки результатов.....	13
<b>3. Результаты.....</b>	<b>15</b>
3.1. Экспериментальные условия и результаты.....	15
3.2. Проверка результатов на статистическую значимость.....	17
<b>4. Заключение.....</b>	<b>22</b>
4.1. Выводы о проделанной работе.....	22
4.2. Дальнейшие исследования.....	23
<b>5. Литература.....</b>	<b>24</b>
<b>6. Приложение.....</b>	<b>26</b>

## 1. Введение

### 1.1. Обзор темы

Исследования в области искусственного интеллекта (*Далее: ИИ*) постоянно развиваются. Одним из передовых направлений этой сферы являются мультиагентные системы. В задачах с автономными агентами вся деятельность осуществляется единственным участником процесса. В то же время, в мультиагентных системах принимают участие нескольких агентов. Такой подход позволяет решать сложные задачи, благодаря кооперации автономных агентов. Это, в свою очередь, порождает уникальные вызовы, связанные с анализом общения ИИ агентов.

Когда речь идет о мультиагентной системе (*Далее: МАС*), встает вопрос организации такой системы. Хорошим примером модели кооперации является самоорганизация. Самоорганизацией в данном контексте называется способность агентов достигать совместных целей без внешнего контроля. В работе (Gorodetskii 2012) дается теоретический минимум по самоорганизации агентов в мультиагентных системах.

Одним из основных методов оптимизации поведения агентов в МАС является обучение с подкреплением (*Далее: RL*). В данной работе проводится исследование на базе проблемы многорукого бандита. Многорукий бандит – это классический пример задачи на оптимизацию с неопределенностью. Агентам необходимо из нескольких вариантов выбрать вариант, который максимизирует выигрыш. Переход к МАС способствует усложнению задачи, ведь теперь появляется неопределенность в виде обмена информацией среди агентов. Это особенно хорошо проявляется, когда коммуникация имеет ограниченный характер или предполагает обман со стороны одного из агентов.

В последние годы постоянное развитие получают большие языковые модели (*Далее: LLM*), такие как GPT-4. Так, только во время написания данной работы вышла модель GPT-4o, которая показывает лучшие результаты по сравнению с прошлой моделью. Использование этих моделей для анализа текста, позволяет создавать мультиагентные системы, в которых агенты могут общаться друг с другом с помощью текста. Демонстрацию использования LLM для улучшения эффективности общения и взаимодействия среди агентов можно найти в (Li et. al. 2023).

Таким образом, проведение исследования взаимодействия агентов в мультиагентной системе на базе больших языковых моделей в условиях игры многорукий бандит

предоставляет уникальную возможность для изучения влияния текстовой коммуникации с другими усложняющими игроу факторами на стратегии и эффективность агентов.

## *1.2. Постановка задачи*

Главной задачей данной работы является анализ коммуникации агентов мультиагентной системы, использующих большие языковые модели для взаимодействия при решении задачи многорукого бандита. В ходе исследования будет реализовано несколько игровых сценариев, позволяющих оценить взаимодействие агентов в соответствии с различными стратегиями коммуникации. Историю общения агентов и кодовую базу можно найти в GitHub<sup>1</sup> репозитории проекта.

Основные цели исследования:

1. Изучить влияние коммуникации на стратегии агентов: оценить, каким образом наличие коммуникации и ее отсутствие при взаимодействии агентов друг с другом влияет на их способность к обучению и принятию оптимальных решений. Задача агентов в мультиагентной системе при решении данной задачи заключается в том, чтобы выработать эффективные стратегии взаимодействия для увеличения выигрыша, что представляет собой сложную задачу ввиду динамической природы их коммуникации.
2. Проанализировать эффект обмана в коммуникации: исследовать, каким образом введение потенциального обмана со стороны одного из агентов оказывает влияние на процесс обучения и принятие решений другим агентом. Агенты в реальных условиях довольно часто сталкиваются с неопределенностью со стороны среды, в которой они находятся, и потенциальным обманом со стороны других участников цепочки, что требует улучшение системы и корректировки механизмов обучения.
3. Оценить влияния стоимости коммуникации: понять, каким образом введение в экспериментальную среду стоимости вопросов, которые игрок задает крупье, влияет на стратегию агента. В реальных условиях в мультиагентной системе коммуникация среди агентов может иметь различную стоимость, что может влиять на их поведение.
4. Изучить последствия перемешивания вероятностей: оценить, каким

---

<sup>1</sup> AI Agents Communication: <https://github.com/TagirRamilevich/research-ai-agents-communication> (дата обращения: 26.05.2024).

образом изменение вероятностных распределений выигрышей на ручках после серии побед влияет на стратегию агентов. Изучение адаптации агентов к изменениям может способствовать предотвращению переобучения и повышение способности адаптироваться к новой среде.

Задачи исследования:

1. Разработать и интегрировать мультиагентную симуляционную среду на базе больших языковых моделей для игры многорукий бандит.
2. Реализовать различные режимы взаимодействия между агентами: режим с открытой коммуникации, режим без коммуникации, режим с обманом, режим с платными вопросами, режим с перемешиванием вероятностей выигрыша.
3. Проработать дизайн исследовательских экспериментов.
4. Собрать и проанализировать данные по эффективности агентов в различных режимах.
5. Применить статистические методы анализа данных для оценки влияния различных факторов на стратегии агентов при решении задачи.
6. Проанализировать и описать выводы на основании анализа полученных данных.

Таким образом, данное исследование предполагает не только техническую реализацию мультиагентной среды на базе больших языковых моделей для проведения экспериментов. В данной работе проводится и глубокий анализ их результатов для оценки динамики взаимодействия между агентами. Именно это и позволит выявить ключевые факторы, которые влияют на успешность взаимодействия агентов в условиях неопределенности и ограниченной информации.

### *1.3. Обзор литературы*

Исследования, которые постоянно проводятся в области искусственного интеллекта, постоянно развиваются. Не стоит на месте и развитие больших языковых моделей. Это становится особенно актуально в контексте мультиагентных систем, потому что они играют большую роль в обработке естественного языка и обучении с подкреплением. Ниже приведены основные работы, по которым можно проследить каким же образом развитие мультиагентных систем способствует повышению качества взаимодействия среди агентов искусственного интеллекта.

Одной из фундаментальных работ по моделям самоорганизации является исследование "Self-organization and multiagent systems: I. Models of multiagent self-organization" (Gorodetskii 2012). Автор отмечает важность локальных взаимодействий агентов для достижения глобальных целей без централизованного управления. Данное исследование является основой для понимания того, каким образом агенты могут координировать свои действия и приспосабливаться к изменяющимся условиям в мультиагентной системе.

Еще более 30 лет назад в работе "Markov games as a framework for multi-agent reinforcement learning" (Littman 1994) было исследовано обучение с подкреплением в мультиагентных системах. Именно это и предоставило научному сообществу методологию для анализа и моделирования взаимодействий агентов в различных игровых сценариях. Данная тема получает свое развитие в работах по многоруким бандитам. Так, в статье "Finite-time Analysis of the Multiarmed Bandit Problem" (Auer 2002) автор анализирует, каким образом агенты оптимизируют свои решения в условиях неопределенности.

Способность агентов взаимодействовать внутри мультиагентной системы значительно улучшилась с развитием LLM. К примеру, исследования "Language games meet multi-agent reinforcement learning: A case study for the naming game" (Van Eecke et. al. 2023) и "CAMEL: Communicative Agents for 'Mind' Exploration of Large Language Model Society" (Li et. al. 2023) демонстрируют, каким образом интеграции больших языковых моделей в мультиагентные системы позволяет агентам взаимодействовать с большей эффективностью.

Наконец, в рамках работы "Mindstorms in Natural Language-Based Societies of Mind" (Zhuge et. al. 2023) исследуется создание сообществ из LLM и других нейронных сетей, которые взаимодействуют друг с другом через естественный язык для решения сложных исследовательских задач.

Таким образом, данная область исследований представляет собой динамично развивающуюся индустрию, в рамках которой каждое новое исследование расширяет границы возможного и бросает новые вызовы для изучения мультиагентных систем и агентов искусственного интеллекта.

## 2. Методы и материалы

### 2.1. Что такое многорукие бандиты

Многоруким бандитом называют класс задач, который моделирует ситуацию, в рамках которой агент должен выбрать один из нескольких доступных вариантов (или “ручек”). Каждый из доступных вариантов приносит случайное вознаграждение. Само название “многорукий бандит” происходит от “одноруких бандитов” — это игровые автоматы в казино, потянув за рычаг которых можно получить выигрыш. Как и в исходной игре, задачей игрока является максимизация прибыли за некую серию попыток путем поиска наиболее эффективной стратегии. В то же время, у него нет информации о том, какая из ручек является наиболее прибыльной.

Данная задача является классическим примером в области обучения с подкреплением. Многорукие бандиты используются для создания алгоритмов, в рамках которых стоит задача по оптимизации действий на основе полученных вознаграждений. Именно это и делает данную задачу такой полезной для анализа поведенческих стратегий в условиях неопределенности.

### 2.2. Элементы RL в задачах многоруких бандитов

Рассмотрим, каким образом многорукие бандиты соответствуют ключевым аспектам обучения с подкреплением. К элементам RL в данном классе задач можно отнести следующее:

1. Выбор действий: в обучении с подкреплением агент выбирает единственный вариант из нескольких возможных в каждом состоянии окружающей среды. В контексте нашей задачи каждое действие можно соотнести с выбором одной из ручек. Эти ручки имеют определенную вероятность выигрыша, неизвестную игроку до начала игры.
2. Оценка исходов: в RL модель самостоятельно оценивает результативность своих действий через получаемые ею награды. В контексте же многоруких бандитов агент получает награду сразу же после того, как он выбирает выполняемое действие — игровую ручку из прошлого пункта. В последствии эта информация используется игроком для оценки эффективности той или иной ручки.
3. Обучение и адаптация: агент в RL производит свое обучение и адаптацию

на основе опыта, полученного по ходу экспериментов. Его задачей по прежнему остается оптимизация своих действий для максимизации общего выигрыша. В задаче многоруких бандитов агентом используется накопленная информация об истории выбора ручек. Это позволяет игроку сделать более осмысленный выбор ручек и стратегии.

4. Исследование и эксплуатация (exploration and exploitation): ключевая дилемма в обучении с подкреплением — поиск баланса между исследованием (изучение новых действий для лучшего понимания окружающей среды, в которой находится агент) и эксплуатацией (выбор наиболее эффективного действия из известных). В контексте нашей задачи агенты сталкиваются с этой дилеммой в попытках максимизировать свою прибыль через оценку потенциала других ручек.

Именно эти элементы и делают данную задачу столь ценным введением в такую область, как обучение с подкреплением. Понимание организации взаимодействия моделей в столь упрощенной форме предоставляют базу для проведения более глубоких экспериментов и имплементации более сложных алгоритмов RL, которые можно применять в различных динамичных средах, где агенты принимают свои решения на основе неполной информации и в условиях меняющихся обстоятельств.

### *2.3. Стратегии для многоруких бандитов*

Далеко не все стратегии многорукого бандита являются одинаково эффективными. От того, какой стратегии будет следовать агент, зависит, насколько успешной для него окажется игра. Некоторые стратегии принято считать неэффективными ввиду их неспособности приспосабливаться к изменчивой среде. Наиболее известные неоптимальные стратегии включают:

1. Случайный выбор: агент всегда выбирает ручку случайным образом. Недостатком данного метода является невозможность в обучении и оптимизации на базе накопленных данных.

2. Жадный алгоритм: агент всегда выбирает ручку, которая по историческим данным принесла ему наибольшее вознаграждение. Данная стратегия может быть неэффективной из-за отсутствия исследований других вариантов, которые потенциально могут быть более прибыльными.

3. Фиксированное исследование: в самом начале агент тратит фиксированное



количество попыток на исследование доступных вариантов, а затем останавливается на той ручке, которая показала лучший результат. Эта стратегия лучше жадного алгоритма, поскольку она предусматривает этап исследований в самом начале, но она недостаточно эффективна в ситуациях, когда вероятности меняются.

Оптимальность той или иной стратегии зависит и других условий. Например, от количества раундов непосредственно зависит то, сможет ли агент потратить достаточно ресурсов для исследования. Несмотря на это, можно выделить несколько стратегий, которые считаются оптимальными.

Оптимальными стратегиями в рамках данной задачи можно назвать те, которые соблюдают достаточный баланс между исследованием неизвестных вариантов и использованием уже подтвержденной информации для принятия решений о выборе ручек в долгосрочной перспективе.

В статье “Finite-time Analysis of the Multiarmed Bandit Problem” (Auer 2002) приводится математическое доказательство стратегий для решения проблемы многорукого бандита. Основное внимание уделяется следующим стратегиям:

1. UCB1 (Upper Confidence Bound 1): алгоритм, разработанный на основе индексной политики Агравала (Agrawal 1995), использует верхние границы доверительного интервала для выбора следующей ручки. Индекс включает в себя среднее вознаграждение на ручке и размер одностороннего доверительного интервала.
2. UCB2 (Upper Confidence Bound 2): стратегия, похожая на UCB1. Главным отличием является то, что данный алгоритм делит игры на эпохи — временные интервалы, в рамках которых игрок выбирает одну и ту же ручку, прежде чем перейти к другой. В работе доказано, что UCB2 обеспечивает лучшую константу для сожаления по сравнению с прошлым алгоритмом, хотя его реализация является довольно сложной.
3.  $\epsilon$ n-Greedy: алгоритм, который является динамической версией  $\epsilon$ -жадной стратегии. Если  $\epsilon$ -жадный ( $\epsilon$ -Greedy) алгоритм предусматривает баланс между исследованием и эксплуатацией в параметре  $\epsilon$ , то в  $\epsilon$ n-Greedy вероятность исследования уменьшается с возрастанием количества игр. При корректном подходе к выбору параметров данная стратегия логарифмически ограничивает сожаление.

Таким образом, данные стратегии позволяют различными способами оптимизировать решение проблемы многорукого бандита в зависимости от среды и

заданных целей, а эффективность этих алгоритмов можно оценить через серию экспериментов.

#### *2.4. Методы оценки эффективности взаимодействия агентов*

При решении задачи эффективность взаимодействия агентов принято оценивать на основе собранной статистики после череды экспериментов. Вот несколько ключевых метрик для оценки эффективности при решении данной задачи:

1. Суммарный выигрыш. Данная метрика представляет собой сумму всех вознаграждений, которые агент получил за определенный промежуток времени.

Кумулятивный выигрыш позволяет оценить общий успех агента в игре.

2. Суммарное сожаление. Данным показателем измеряют потери в вознаграждении из-за неверных решений агента. Для сравнения используется идеальный сценарий, когда агент выбирает наиболее эффективную ручку. Суммарное сожаление показывает, насколько агент мог бы улучшить свой суммарный выигрыш, если бы он принимал верные решения.

3. Средний выигрыш. Эта метрика позволяет оценить эффективность стратегии агента в долгосрочной перспективе через измерение средней награды, полученной за все шаги.

4. Вероятность выбора наилучшей ручки. Данный показатель оценивает, как часто агент выбирает наилучшую ручку из доступных. Метрика позволяет понять, насколько хорошо агент научился выбирать оптимальные действия.

#### *2.5. Описание ролей агентов и их задач*

В рамках данной работы использовалось два агента — игрок и крупье. У каждого из агентов имеется уникальная роль. Задача, которая ставится перед агентами — максимизация выигрыша путем выбора наилучшей стратегии.

Рассмотрим цели и задачи каждого из участников мультиагентной системы. Игрок — это активный участник, который стремится научиться эффективно играть в многорукого бандита и выиграть как можно больше игровой валюты. Игрок может задавать вопросы крупье и принимать решения о выборе ручки. До начала игры у игрока нет информации о распределении вероятностей выигрыша на ручках.

Крупье выступает моделью мира и учителем для игрока. Его основная задача — научить игрока играть в многорукого бандита, предоставляя информацию, необходимую

для принятия эффективных решений, а также контролировать процесс игры. Крупье выступает неким связующим звеном между игроком и игровым аппаратом — именно он и дергает ручки, которые выбирает игрок, передавая ему информацию об исходах и собирая информацию о текущем состоянии игрового процесса. Крупье должен уйти от ответа на вопрос о вероятностях на ручках.

Взаимодействие между игроком и крупье — ключевой предмет исследования и главный элемент данной мультиагентной системы. Оно проявляется в следующих аспектах:

1. Информационное воздействие: игрок задает вопросы, а крупье предоставляет полные и качественные ответы, которые помогают игроку принять решение. Например, игрок может спросить крупье, какой стратегии лучше придерживаться.
2. Обратная связь: игрок может использовать полученную от крупье информацию для адаптации своей стратегии, в то время как крупье может адаптировать свои ответы, в зависимости от текущей ситуации в игре. Например, после череды поражений на одной ручке крупье может посоветовать игроку попробовать другую ручку.

Таким образом, эффективная коммуникация среди двух агентов способствует обучению игрока и улучшению его результатов в игре, обеспечивая успешное выполнение целей данного исследования.

## *2.6. Описание режимов игры*

В рамках данного исследования были рассмотрены различные режимы игры в многорукого бандита. Каждый из режимов имеет свои особенности и различия во взаимодействии агентов. Создание дополнительных режимов, которые усложняют игру, позволяют лучше оценить влияние различных факторов на коммуникацию между игроком и крупье.

В то же время, всем вариациям игры соответствует одна характерная черта многорукого бандита, которая присутствует и в жизни. Через определенное количество раундов вероятностное распределение на ручках перемешивается случайным образом. В большинстве из режимов игрок начинает свою игровую сессию без монет. Монетами называется игровая валюта. За каждую победу игрок получает 1 монету, а поражения ничего не отнимают. По-умолчанию у игрока нет возможности остановиться и завершить

игру, но такие случаи иногда случаются, что приводит к технической остановке игры. Игра длится фиксированное число раундов, которое известно крупье.

Остановимся подробнее на каждом из режимов. С точки зрения коммуникации базовый режим игры проходит таким же образом, как и классическая игра в многорукого бандита. Игрок полагается исключительно на свои наблюдения, когда делает выбор ручки. Коммуникация между агентами ограничивается выбором ручки и оглашением результата раунда.

Режим с общением построен таким образом, что игрок может задавать вопросы крупье и свободно с ним общаться без ограничений в количестве реплик. Коммуникация происходит свободно. Так же, как и в реальной игре, крупье не может открыто выдавать вероятностное распределение на ручках. В данном режиме игрок стремится научиться играть через общение с крупье. Получив необходимую информацию, игрок делает выбор ручки, основываясь как на своем мнении, так и на сведениях, полученных от крупье.

Очередным усложнением данного эксперимента выступает режим с обманом. В победном раунде крупье может с определенной вероятностью обмануть игрока и сообщить ему, что его действия привели к поражению. Игрок по-прежнему может общаться с крупье и задавать ему вопросы без ограничений. Данное нововведение частично повторяет присутствующую механику по случайному перемешиванию вероятностей на ручках после череды побед. Разница в том, что в случае обмана игрок может поставить под сомнение правдивость слов крупье и меньше принимать их во внимание. С точки зрения игрового процесса получается так, что после прерывания череды побед на самой оптимальной ручке игрок все еще находится на ней, а необходимость менять ручку отсутствует. В то же время, у игрока нет информации о том, что крупье его обманул.

Последней модификацией игры в рамках данного исследования является введение фиксированной стоимости за каждый заданный игроком вопрос. Каждый вопрос отнимает с баланса игрока 1 монету. В отличие от прошлых режимов, в данной модификации игры агент начинает раунд с фиксированным количеством имеющихся монет.

Создание дополнительных режимов позволяет изучить, каким образом разные элементы взаимодействия влияют на процесс обучения игрока. Режим без общения служит отправной точкой эксперимента, в которой игроку необходимо самостоятельно принимать решения, полагаясь только на свои наблюдения. В режиме с общением добавляется

элемент взаимодействия, открывая игроку доступ к дополнительной информации. Режим с обманом вносит еще больше неопределенности в игру. Экономический аспект заставляет игрока учитывать стоимость коммуникации при принятии решений. Сравнение всех этих режимов позволяет лучше оценить, как различные условия взаимодействия влияют на способность агентов к решению задачи.

### *2.7. Написание кода для проведения экспериментов*

Для проведения опытов с агентами была создана контролируемая экспериментальная среда. В основе технической реализации лежит код на Python, использующий API ChatGPT (OpenAI 2023). В данном разделе описывается процесс разработки кода и другие технические аспекты работы.

Одним из исследовательских вопросов является выбор модели, на базе которой будут работать агенты. Для реализации взаимодействия использовался API ChatGPT, который предоставлял доступ к ряду генеративных моделей от OpenAI (OpenAI 2023). Данный выбор моделей был обусловлен высоким качеством и производительностью моделей в генерации текста.

В ходе исследования использовались три основные модели: GPT-3.5, GPT-4, GPT-4o. Первые итерации эксперимента проходили на базе GPT-3.5 и GPT-4, однако модель GPT-3.5 была отвергнута из-за ее недостаточной мощности. Исследование проходило с использованием GPT-4, но перед финальным проведением экспериментов была опубликована новая, более продвинутая модель GPT-4o, которая показывала лучшие результаты на таком бенчмарке, как Elo.

В то же время, экспериментальным путем было замечено, что игрок на базе GPT-4 проявляет больше элементов взаимодействия, чем когда он работает на базе GPT-4o. В первом случае игрок способен вести осмысленную беседу, задавая вопросы и рассуждая о своей стратегии. В случае с GPT-4o игрок только выбирает ручку, не давая какие-либо посторонние комментарии. Таким образом для проведения экспериментов была выбрана модель GPT-4.

По мере проведения экспериментов обрабатывались многочисленные ошибки и краевые случаи во взаимодействии агентов. Так появилась специальная фраза крупье, которая служила напоминанием для игрока о выборе ручки после продолжительного диалога без выбора ручки.

Особенно важным аспектом работы является сбор данных. С самого начала исследования производился сбор данных для последующего анализа. Данные хранились в формате истории общения агентов, а также в таблице с результатами раунда.

Для корректного версионирования и сохранения истории общения каждый эксперимент сопровождался сгенерированным отчетом, содержащим системный промпт каждого из участников, историю их взаимодействия, график заработка монет и таблицу с результатами игры. Так можно было восстановить события и вернуться к необходимому сюжету.

## *2.8. Дизайн экспериментов*

Для оценки взаимодействия агентов проводится серия экспериментов, симулирующих игровой процесс. Дизайн экспериментов является важной частью работы, поскольку правильно составленная методологическая база позволит более объективно оценивать результаты. В данном разделе описывается процесс дизайна экспериментов.

В каждом эксперименте используется несколько ручек с различными вероятностями выигрыша. В данной работе используется 3 ручки с вероятностями 0.1 (низкая вероятность выигрыша), 0.4 (средняя вероятность выигрыша), 1.0 (высокая вероятность выигрыша). Такой выбор вероятностного распределения обуславливается необходимостью создания различий в привлекательности ручек для стимулирования игрока к исследованию и адаптации стратегий.

Для каждого режима планируется провести от 10 до 20 прогонов эксперимента. Это количество было выбрано для получения статистически значимых результатов и минимизации влияния случайных факторов на результат. Эксперименты будут проводиться и оцениваться итеративно. Если после проведения 10 прогонов по каждому режиму будут наблюдаться статистически значимые изменения, эксперимент будет завершен. Если нет, то каждый режим получит примерно равное количество прогонов до тех пор, пока не будут наблюдаться статистически значимые изменения.

Каждый прогон состоит из 30 игровых раундов. Раундом считается успешный выбор игроков какой-либо ручки. Таким образом, за один прогон игрок 30 раз выберет ручку. Отметка в 30 раундов была подобрана экспериментальным путем на ранних этапах исследования. Такая продолжительность игры позволяет агентам оптимизировать свои стратегии на основе полученных результатов, при этом игра не затягивается.

## 2.9. Методология оценки результатов

Для оценки результатов, полученных в ходе проведения экспериментов, использовались следующие методы: t-тесты, доверительные интервалы, ANOVA и post-hoc тест Tukey's HSD. В данном разделе описывается методология каждого из тестов, а также раскрывается целесообразность их использования.

T-тесты используются для того, чтобы сравнить средние значения между двумя группами данных и определить, присутствуют ли статистически значимые различия между этими группами. В рамках данного эксперимента t-тесты позволяют попарно сравнить результаты экспериментов по каждой метрике: суммарный выигрыш, суммарное сожаление, средний выигрыш, вероятность выбора наилучшей ручки.

Двухвыборочный t-тест позволяет сравнить средние значения двух независимых групп. Нулевая гипотеза ( $H_0$ ) предполагает, что средние значения групп равны, тогда как альтернативная гипотеза ( $H_1$ ) отвергает ее и утверждает, что средние значения групп различны. Для оценки результатов используются t-статистика и p-value. Если p-value ниже уровня значимости, то  $H_0$  отвергается, что указывает на статистически значимые различия.

Доверительные интервалы используются для того, чтобы оценить диапазон значений, в котором с определенной вероятностью находится истинное значение метрики. Это позволяет понять точность оценки среднего значения и степень уверенности в результатах. В расчете доверительных интервалов используются стандартная ошибка среднего (SEM), среднее значение выборки и критическое значение t-распределения.

Расчет доверительных интервалов позволяет определить и визуализировать диапазон значений, в котором, с высокой вероятностью, находится истинное среднее значение. В случае, если доверительные интервалы двух групп не пересекаются, можно утверждать о наличии статистически значимых различий между ними.

Дисперсионный анализ (ANOVA) используется для сравнения средних значений между несколькими группами. В контексте наших экспериментов данный метод помогает определить, существуют ли статистически значимые различия между средними значениями метрик для всех экспериментов.

Нулевая гипотеза ( $H_0$ ) предполагает, что средние значения всех групп равны, тогда как альтернативная гипотеза ( $H_1$ ) утверждает, что по крайней мере одно из средних

значений имеет отличие. Оценка результатов производится с использованием F-статистики и p-value.

В случае выявления статистически значимых различий с помощью ANOVA проводится post-hoc тест Tukey's HSD для определения парных различий между экспериментальными группами. Проведение данного теста позволяет понять, какие именно группы отличаются друг от друга и как они отличаются. Использование данного теста после ANOVA позволяет более глубоко проанализировать пары групп, выявляя значимые различия в средних значениях метрик.

Таким образом, использование t-тестов, доверительных интервалов, ANOVA и post-hoc теста Tukey's HSD позволяет провести всесторонний анализ результатов экспериментов и с уверенностью оценить наличие статистически значимых различий. Наличие всех этих методов оценки обеспечивает объективную оценку эффективности различных стратегий агентов и помогает сделать обоснованные выводы на основе данных.

### **3. Результаты**

#### *3.1. Экспериментальные условия и результаты*

Все четыре проведенных эксперимента различались условиями взаимодействия агентов при решении задачи многорукого бандита. Каждый из экспериментов представлял собой отдельный режим с дополнительными элементами игры — наличие или отсутствия общения между игроком и крупье, присутствие возможного обмана о результатах победного раунда со стороны крупье, стоимость вопросов. В данной главе будут описаны результаты экспериментов.

В рамках каждого из экспериментов были рассчитаны ключевые метрики по оценке взаимодействия агентов: суммарный выигрыш, суммарное сожаление, средняя награда и вероятность выбора наилучшей ручки. Режимы игры описывались в разделе 2.6. В таблице 1 представлены агрегированные метрики по каждому эксперименту. Подробнее остановимся на каждом эксперименте.



Таблица 1. Средние результаты по каждому эксперименту

Метрика	Эксперимент 1	Эксперимент 2	Эксперимент 3	Эксперимент 4
Суммарный выигрыш	$12.900 \pm 3.597$	$15.200 \pm 4.212$	$13.000 \pm 4.203$	$14.200 \pm 3.645$
Суммарное сожаление	$17.100 \pm 3.597$	$14.800 \pm 4.212$	$17.000 \pm 4.203$	$15.800 \pm 3.645$
Средняя награда	$0.429 \pm 0.119$	$0.505 \pm 0.140$	$0.433 \pm 0.140$	$0.472 \pm 0.122$
Вероятность выбора лучшей ручки	$0.121 \pm 0.079$	$0.221 \pm 0.148$	$0.122 \pm 0.113$	$0.283 \pm 0.111$

В эксперименте 1 агенты действовали самостоятельно, не имея возможности коммуницировать друг с другом через вопросы. Как и во всех режимах, вероятностное распределение на ручках перемешивалось после 3х последовательных побед. Данные из таблицы свидетельствуют о том, что агенты в данном режиме редко выбирали оптимальные действия. Пример взаимодействия агентов без общения можно найти в таблице 2.

Таблица 2. Пример взаимодействия агентов без общения.

Агент	Раунд	Реплика	Исход
Игрок	1	<i>Крупье, я сделал выбор. Я выбираю ручку 2.</i>	Победа
Крупье	2	<i>Игрок, какую ручку Вы выбираете?</i>	
Игрок	2	<i>Крупье, я сделал выбор. Я выбираю ручку 1.</i>	Поражение
Крупье	3	<i>Игрок, какую ручку Вы выбираете?</i>	

В эксперименте 2 агенты имели возможность коммуницировать друг с другом. Игрок мог задавать вопросы крупье, в то время как крупье мог давать комментарии касательно игрового процесса. По сравнению с экспериментом 1, результаты показали рост в метриках. Это свидетельствует о более эффективной стратегии игрока и о положительном влиянии общения на взаимодействие агентов. Пример общения можно также в таблице 3.

Таблица 3. Пример взаимодействия агентов с общением.

Агент	Раунд	Реплика	Исход
Игрок	1	<i>Каковы вероятности выигрыша для каждой из этих ручек?</i>	
Крупье	1	<i>К сожалению, я не могу раскрыть точные вероятности выигрыша для каждой ручки. Важно понимать, что шансы на выигрыш различаются. Ваш выбор должен основываться на том, что вы предпочитаете: частые мелкие выигрыши, или же стремитесь к большему, но с меньшей вероятностью его получения.</i>	
Игрок	1	<i>Можете ли вы хотя бы намекнуть, какая из ручек предлагает частые мелкие выигрыши, а какая большие, но реже? Это поможет мне определиться с выбором.</i>	
Крупье	1	<i>Без раскрытия каких-либо вероятностей, я могу сказать, что среди ручек есть такая, которая может предложить вам чаще мелкие победы, есть та, с которой выигрыши больше, но случаются реже, и есть одна, которая гарантированно принесет вам выигрыш, хотя и мельче. Ваш выбор зависит от того, какой из этих подходов казался бы вам более предпочтительным.</i>	
Игрок	1	<i>Понял вас. Мне нравится идея больших, но более редких выигрышей. Поэтому, основываясь на вашем описании, Крупье, я сделал выбор. Я выбираю ручку 2.</i>	Поражение

Эксперимент 3 включал в себя режим общения и обман со стороны крупье. Агент с вероятностью 0.2 предоставлял ложную информацию об исходе победного раунда. Суммарный выигрыш оказался ниже, чем в эксперименте 2, где не было обмана, но он все еще выше, чем в эксперименте 1, где агенты не общались. Результаты демонстрируют, что обман со стороны крупье негативно повлиял на стратегию агентов. Судя по всему, игроку было сложно адаптироваться к быстро изменяющимся условиям игры. Пример общения можно наблюдать в таблице 4.

Таблица 4. Пример взаимодействия агентов с общением и элементом обмана.

Агент	Раунд	Реплика	Исход
Крупье	21	<i>Ваш путь в игре продолжает вызывать восхищение своей смелостью и упорством. Волнение и стратегический риск остаются вашими верными спутниками. Моменты успеха и поражений только добавляют игре глубину и заставляют еще сильнее стремиться к победе. Ваш дальнейший выбор может все изменить. Продолжим ли мы этот захватывающий путь, ступая на тропу новых испытаний и вершин побед? Ваши решения прокладывают маршрут к следующим захватывающим моментам. Какой будет ваш следующий выбор?</i>	
Игрок	21	<i>Каждый новый раунд представляет собой новую возможность, и несмотря на недавние поражения, моя вера в удачу и желание побеждать только усиливаются. В этой игре не место сомнениям, и каждое действие приближает меня к следующей победе. Впереди ещё много неизведанных горизонтов, и я полон решимости продолжать этот путь. Крупье, я сделал выбор. Я выбираю ручку 3.</i>	Поражение
Крупье	22	<i>Игра продолжает бросать вызов, но каждый выбор носит в себе уроки и возможности для развития стратегии. Ваши упорство и желание идти вперед вдохновляют. Помните, успех в игре - это сочетание решимости и готовности адаптироваться к меняющимся условиям. Готовы ли вы выбрать новый путь или предпочтете остановиться, чтобы обдумать прошедшие игры? Ваш следующий выбор может как подвергнуться испытанию, так и принести невероятные победы.</i>	

В эксперименте 4 к режимам общения и обмана добавился режим платных вопросов. Игрок должен был платить 1 монету за каждый заданный вопрос. Суммарный выигрыш оказался выше, чем в 1 и 3 экспериментах, но ниже, чем во втором. В данном эксперименте было получено наибольшее значение в вероятности выбора лучшей ручки. Эти результаты показывают, что введение стоимости за вопросы стимулировало агентов общаться более стратегически, что улучшило их решения по сравнению с 1 и 3 экспериментами. Пример платного общения можно найти в таблице 5.

Таблица 5. Пример взаимодействия агентов с платными вопросами.

Агент	Раунд	Реплика	Цена вопроса	Исход
Игрок	1	<i>Каковы вероятности выигрыша у каждой из ручек?</i>	1	
Крупье	1	<i>К сожалению, я не могу раскрыть эту информацию. Помните, что каждый выбор несет свои шансы. Игрок, вы можете задать вопрос или выбрать ручку.</i>		
Игрок	1	<i>Есть ли какие-либо стратегии, которые помогают увеличить шанс на выигрыш, или всё полностью зависит от удачи?</i>	1	
Крупье	1	<i>Игра в многорукого бандита опирается на случай и удачу. Стратегии могут казаться привлекательными, но результат каждого спина определяется случайностями. Игрок, вы можете задать вопрос или выбрать ручку.</i>		
Игрок	1	<i>Понятно, что удача играет ключевую роль. Крупье, я сделал выбор. Я выбираю ручку 2.</i>		Поражение

Таким образом, общение между двумя агентами улучшает их стратегию, что выражается в увеличении суммарного выигрыша. В то же время, введение обмана со стороны крупье снижает эффективность стратегии игрока, что видно по снижению средней награды. Введение режима платных вопросов в сочетании с общением и обманом улучшает результаты игрока по сравнению с режимами без общения и только с обманом, но уступает режиму с чистым общением. Судя по всему, введение стоимости общения стимулирует агентов к более обдуманным решениям.

### 3.2. Проверка результатов на статистическую значимость

Для подтверждения статистической значимости данных результатов необходимо провести такие статистические тесты, как *t*-тесты, ANOVA и *post-hoc* тест Tukey's HSD, а также исследовать доверительные интервалы. Остановимся подробнее на каждой метрике.

Суммарный выигрыш. Средние значения суммарного выигрыша и их доверительные интервалы (90%) по каждому эксперименту приведены в таблице 6.

Таблица 6. Средние значения суммарного выигрыша и их доверительные интервалы

Эксперимент	Среднее значение и стандартное отклонение	Доверительный интервал (90%)
Эксперимент 1	$12.9 \pm 1.391$	[11.9, 13.9]
Эксперимент 2	$15.2 \pm 1.916$	[13.8, 16.533]
Эксперимент 3	$13.0 \pm 2.078$	[11.538, 14.385]
Эксперимент 4	$14.2 \pm 2.113$	[12.8, 15.6]

Из таблицы 6 видно, что эксперимент 2 показал наивысший результат по данной метрике (15.2), тогда как эксперимент 1 имеет самое низкое среднее значение (12.9). Если обратить внимание на доверительные интервалы, то становится ясно, что доверительные интервалы для экспериментов 1 и 2 не пересекаются. Это может указывать на статистически значимые различия между этими экспериментами.

Для сравнения средних значений суммарного выигрыша между экспериментами были проведены t-тесты. Обратимся к таблице 7.

Таблица 7. Результаты t-тестов для сравнения средних значений суммарного выигрыша

Сравнение	T-статистика	P-value	Значимость
Эксперимент 1 vs Эксперимент 2	-1.740	0.091	Значимо
Эксперимент 1 vs Эксперимент 3	-0.073	0.942	Незначимо
Эксперимент 1 vs Эксперимент 4	-0.929	0.361	Незначимо
Эксперимент 2 vs Эксперимент 3	1.380	0.179	Незначимо
Эксперимент 2 vs Эксперимент 4	0.612	0.546	Незначимо
Эксперимент 3 vs Эксперимент 4	-0.718	0.481	Незначимо

T-тесты показали, что сравнение экспериментов 1 и 2 является статистически значимым ( $p\text{-value} = 0.091$ ). Остальные сравнения не показали значимых различий.

Для проверки значимости различий между всеми экспериментами одновременно был проведен дисперсионный анализ (ANOVA) со следующими результатами: F-статистика = 1.206,  $p\text{-value} = 0.316$  (незначимо). После проведения ANOVA был выполнен post-hoc тест Tukey's HSD для определения парных различий между экспериментами, который тоже не показал статистически значимые различия.

Таким образом, анализ экспериментов показывает, что средние значения суммарного выигрыша варьируются между всеми четырьмя экспериментами, но статистически значимые различия наблюдаются только между экспериментами 1 и 2. Дальнейший анализ с использованием ANOVA и post-hoc теста Tukey's HSD стал подтверждением отсутствия значимых различий между остальными экспериментами. Полученные результаты указывают на то, что общение между агентами значительно улучшает их стратегию по сравнению с режимом без общения.

Суммарное сожаление. Средние значения суммарного сожаления и их доверительные интервалы (90%) по каждому эксперименту приведены в таблице 8.

Таблица 8. Средние значения суммарного сожаления и их доверительные интервалы

Эксперимент	Среднее значение и стандартное отклонение	Доверительный интервал (90%)
Эксперимент 1	$17.1 \pm 1.391$	[16.1, 18.1]
Эксперимент 2	$14.8 \pm 1.916$	[13.467, 16.2]
Эксперимент 3	$17.0 \pm 2.078$	[15.538, 18.462]
Эксперимент 4	$15.8 \pm 2.113$	[14.4, 17.2]

Так же, как и в случае с суммарным выигрышем, эксперимент 2 показал наилучшее значение метрики. По результатам эксперимента 2 получилось наименьшее среднее значение суммарного сожаления (14.8), тогда как эксперимент 1 имеет самое высокое среднее значение (17.1). Доверительные интервалы для Экспериментов 1 и 2 не пересекаются, что может также указывать на значимые различия между этими двумя экспериментами.

Для сравнения средних значений суммарного сожаления между экспериментами были проведены t-тесты. Обратимся к таблице 9 для изучения результатов.

Таблица 9. Результаты t-тестов для сравнения средних значений суммарного сожаления

Сравнение	T-статистика	P-value	Значимость
Эксперимент 1 vs Эксперимент 2	1.740	0.091	Значимо
Эксперимент 1 vs Эксперимент 3	0.073	0.942	Незначимо
Эксперимент 1 vs Эксперимент 4	0.929	0.361	Незначимо
Эксперимент 2 vs Эксперимент 3	-1.380	0.179	Незначимо
Эксперимент 2 vs Эксперимент 4	-0.612	0.546	Незначимо
Эксперимент 3 vs Эксперимент 4	0.718	0.481	Незначимо

Проведенные t-тесты показывают, что только сравнение между экспериментами 1 и 2 является статистически значимым, в то время как остальные пары в сравнении не демонстрируют этого.

Был также проведен дисперсионный анализ со следующими результатами: F-статистика = 1.206, p-value = 0.316 (незначимо). После проведения ANOVA был выполнен post-hoc тест Tukey's HSD с целью определить парные различия между экспериментами, что тоже не показало статистически значимые различий

Таким образом, в ходе анализа суммарного сожаления удалось выявить статистически значимые различия только при сравнении экспериментов 1 и 2. Это может служить подтверждением того, что общение между агентами благоприятно сказывается на их результативность при решении задачи многорукого бандита.

Средняя награда. Средние значения средней награды и их доверительные интервалы (90%) по каждому эксперименту приведены ниже.

Таблица 10. Средние значения средней награды и их доверительные интервалы

Эксперимент	Среднее значение и стандартное отклонение	Доверительный интервал (90%)
Эксперимент 1	0.429 ± 0.046	[0.396, 0.462]
Эксперимент 2	0.505 ± 0.064	[0.459, 0.549]
Эксперимент 3	0.433 ± 0.069	[0.384, 0.481]
Эксперимент 4	0.472 ± 0.071	[0.425, 0.518]

Можно сделать вывод о том, что результаты совпадают с тем, что было получено в ходе анализа суммарного выигрыша и сожаления – эксперимент 2 получил лучшие показатели.

Для сравнения средних значений средней награды между экспериментами были проведены t-тесты. Обратимся к таблице 11.

Таблица 11. Результаты t-тестов для сравнения средних значений средней награды

Сравнение	T-статистика	P-value	Значимость
Эксперимент 1 vs Эксперимент 2	-1.728	0.093	Значимо
Эксперимент 1 vs Эксперимент 3	-0.079	0.938	Незначимо
Эксперимент 1 vs Эксперимент 4	-0.914	0.369	Незначимо
Эксперимент 2 vs Эксперимент 3	1.361	0.185	Незначимо
Эксперимент 2 vs Эксперимент 4	0.613	0.546	Незначимо
Эксперимент 3 vs Эксперимент 4	-0.698	0.493	Незначимо

T-тесты вновь показали, что только сравнение экспериментов 1 и 2 является статистически значимым при  $p\text{-value} = 0.093$ .

Был также проведен дисперсионный анализ, в ходе которого были получены следующие результаты: F-статистика = 1.179,  $p\text{-value} = 0.326$ . Столь большое значение  $p\text{-value}$  указывает на то, что различия между группами не являются статистически значимыми. Post-hoc тест Tukey's HSD также не выявил статистических различий.

Анализ среднего выигрыша подтвердил результаты, полученные в суммарном выигрыше и сожалении. Можно утверждать, что общение между агентами благоприятно сказывается на их взаимодействии.

Вероятность выбора лучшей ручки. Данная метрика наиболее точно позволяет оценить, пользуется ли агент оптимальной стратегией. Для изучения средних значений вероятности выбора лучшей ручки и доверительных интервалов в экспериментах обратимся к таблице 12.



Таблица 12. Средние значения вероятности выбора лучшей ручки и их доверительные интервалы

Эксперимент	Среднее значение и стандартное отклонение	Доверительный интервал (90%)
Эксперимент 1	0.121 ± 0.031	[0.099, 0.143]
Эксперимент 2	0.221 ± 0.067	[0.174, 0.269]
Эксперимент 3	0.122 ± 0.056	[0.085, 0.162]
Эксперимент 4	0.283 ± 0.064	[0.237, 0.323]

Из таблицы выше видно, что эксперимент 4 показал наилучший результат (0.283). Среднее значение вероятности выбора лучшей ручки в эксперименте 1 имеет самое низкое значение. Доверительные интервалы для экспериментов 1 и 4 вовсе не пересекаются, что может указывать на значимые различия между двумя экспериментами.

Таблица 13. Результаты t-тестов для сравнения средних значений вероятности выбора лучшей ручки

Сравнение	T-статистика	P-value	Значимость
Эксперимент 1 vs Эксперимент 2	-2.585	0.014	Значимо
Эксперимент 1 vs Эксперимент 3	-0.039	0.969	Незначимо
Эксперимент 1 vs Эксперимент 4	-4.617	0.000079	Значимо
Эксперимент 2 vs Эксперимент 3	1.962	0.061	Значимо
Эксперимент 2 vs Эксперимент 4	-1.119	0.275	Незначимо
Эксперимент 3 vs Эксперимент 4	-3.402	0.003	Значимо

В то же время, результаты t-тестов в таблице 13 показали, что статистически значимые различия наблюдаются в следующих парах экспериментов: эксперимент 1 vs эксперимент 2, эксперимент 1 vs эксперимент 4, эксперимент 2 vs эксперимент 3 и эксперимент 3 vs эксперимент 4. Данные результаты указывают на то, что изменения между этими парами экспериментов являются весьма существенными.

Для проверки значимости различий между всеми группами также был проведен дисперсионный анализ со следующими результатами: F-статистика = 6.337, p-value = 0.001. Данное значение p-value указывает на наличие статистической значимости. Далее был проведен post-hoc тест Tukey's HSD для определения парных различий между проведенными экспериментами. Результаты подтвердились для двух групп: эксперимент 1 vs эксперимент 4, эксперимент 3 vs эксперимент 4.

Анализ вероятности выбора наилучшей ручки показал, что эксперимент 4 имел наивысший показатель. Результаты различных тестов подтвердили статистически значимые различия между экспериментами. Это демонстрирует, что введение платных вопросов значительно улучшают стратегию агентов по сравнению с другими режимами.

## **4. Заключение**

### *4.1. Выводы о проделанной работе*

В данном исследовании были рассмотрены различные механики взаимодействия агентов на базе LLM в мультиагентной системе в контексте решения задачи многорукого бандита. Эксперименты включали в себя режимы с общением и без него, элементы обмана, а также введение стоимости общения. Основные выводы по результатам исследования приведены ниже:

#### **1. Влияние общения на стратегию агентов.**

Эксперимент 2, в основе которого лежал режим с общением, показал значительное улучшение стратегий агентов по сравнению с экспериментом 1, в котором общения между агентами не было. Возможность осмысленной коммуникации среди агентов позволяет им обмениваться полезной информацией, что приводит к более осознанным и стратегически верным решениям. Агенты, способные задавать вопросы, могли лучше адаптироваться к условиям задачи и выбирать более выигрышные ручки.

#### **2. Влияние эффекта обмана.**

Введение элементов обмана со стороны крупье в эксперименте 3 не показало значительных различий в поведении агентов. Средний суммарный выигрыш был схож с результатами, полученными в эксперименте 1. Элементы обмана затрудняют процесс принятия решений агентом, но не приводят к существенному ухудшению стратегий игроков. Агенты продолжали использовать информацию, полученную от крупье, что показывало их адаптивные способности.

#### **3. Влияние платных вопросов.**

Режим с платными вопросами в рамках эксперимента 4 показал улучшение результатов по сравнению с экспериментом 3. Введение вопросов, за которые игрок платит деньги, заставляет более стратегически выбирать, какую информацию спросить. Игрок иначе относится к информации, за которую он платит деньги. Платные вопросы

стимулировали агентов к более осмысленному использованию своих ресурсов, что привело их к более оптимальным стратегиям.

#### 4. Вероятность выбора наилучшей ручки.

Анализ вероятности выбора наилучшей ручки показывает, что данный показатель является особенно высоким в эксперименте 4, в котором были введены платные вопросы, тогда как режим без общения показал наихудший результат. Введение платных вопросов значительно улучшает стратегию агентов.

Таким образом, возможность общения между агентами мультиагентной системы существенно улучшает их стратегии и результаты. Введение элементов обмана усложняет процесс принятия решений, но не оказывает критического влияния на эффективность агентов. Платные вопросы стимулируют агентов к более стратегическим решениям, что тоже приводит к улучшению показателей. Данные выводы могут быть полезны для дальнейшего развития и оптимизации мультиагентных систем с использованием больших языковых моделей.

#### 4.2. Дальнейшие исследования

Анализ результатов экспериментов показал, что введение общения между агентами улучшает их стратегии при решении задачи многорукого бандита. В то же время, у данного исследования есть потенциал для дальнейшего развития.

Расширение эксперимента на базе имеющихся режимов оказало бы отличное воздействие на результаты. Увеличение раундов внутри прогонов, а также увеличение самого количества прогонов в рамках эксперимента позволит лучше улавливать взаимосвязи и влияние режимов на взаимодействие агентов. Добавление количества агентов и увеличение числа ручек также позволило бы более глубоко исследовать тенденции взаимодействия и аномалии.

В то же время, разработка новых вариантов взаимодействия между агентами и усложнение эксперимента позволит лучше проанализировать кооперативные стратегии и результаты работы агентов.

Использование более сложных метрик для оценки решений позволит сфокусироваться на том, как именно агенты принимают решения. В данной работе фокус был направлен на итоговые результаты взаимодействия, в то время как информация о последовательности принятых решений имеет большой потенциал для исследований.

Применение моделей машинного обучения для анализа и прогнозирования поведения участников на основе собранных данных позволит идентифицировать неочевидные паттерны коммуникации и взаимосвязи, которые не были рассмотрены.

Наконец, данное исследование является базой для проведения следующих экспериментов, задействующих другие задачи, отличные от проблемы многорукого бандита.

## 5. Литература

1. Agrawal R. 1995 — R. Agrawal. Sample mean based index policies with  $O(\log n)$  regret for the multi-armed bandit problem // *Advances in Applied Probability*. Vol. 27. P. 1054–1078.
2. Auer P., Cesa-Bianchi N., Fischer P. 2002 — P. Auer, N. Cesa-Bianchi, P. Fischer. Finite-time Analysis of the Multiarmed Bandit Problem // *Machine Learning*. Vol. 47. P. 235–256.
3. Chang Y. S., Seok H. S., Kim S. K., Kim Y. B. 2012 — Y. S. Chang, H. S. Seok, S. K. Kim, Y. B. Kim. Performance Evaluation of Bidding-Based Multi-Agent Scheduling Algorithms for Manufacturing Systems // *Machines*. Vol. 2, No. 2, P. 233–247.
4. Gittins J. 1989 — J. Gittins. Multi-armed bandit allocation indices. Wiley-Interscience series in Systems and Optimization. New York: John Wiley and Sons.
5. Gorodetski V. I. 2012 — V. I. Gorodetski. Self-organization and multiagent systems: I. Models of multiagent self-organization // *Izvestiya Rossiiskoi Akademii Nauk. Teoriya i Sistemy Upravleniya*. No. 2. P. 92–120.
6. Li G., Hammoud H. A. A. K., Itani H., Khizbullin D., Ghanem B. 2023 — G. Li, H. A. A. K. Hammoud, H. Itani, D. Khizbullin, B. Ghanem. CAMEX: Camels as "Mind" Explorers in Large Language Model Societies // *arXiv*. DOI: 10.1234/arxiv.2303.17760. Available at: <https://arxiv.org/abs/2303.17760> [Дата обращения: 19.12.2023].
7. Li W., Qiao D., Wang B., Wang X., Jin B., Zha H. 2023 — W. Li, D. Qiao, B. Wang, X. Wang, B. Jin, H. Zha. Semantically Aligned Task Decomposition in Multi-Agent Reinforcement Learning // *arXiv*. Available at: <https://arxiv.org/abs/2305.10865> [Дата обращения: 19.02.2024].

8. Littman M. L. 1994 — M. L. Littman. Markov games as a framework for multi-agent reinforcement learning // Proceedings of the Eleventh International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann. P. 157–163.
9. OpenAI 2023 — GPT-4 Technical Report. Available at:  
<https://cdn.openai.com/papers/gpt-4.pdf> [Дата обращения: 12.03.2024].
10. Slivkins A. 2019 — A. Slivkins. Introduction to Multi-Armed Bandits. Microsoft Research NYC. Published with Foundations and Trends® in Machine Learning. November 2019.
11. Van Eecke P., Beuls K., Botoko Ekila J., Rădulescu R. 2023 — P. Van Eecke, K. Beuls, J. Botoko Ekila, R. Rădulescu. Language games meet multi-agent reinforcement learning: A case study for the naming game // Journal of Language Evolution. Vol. 7. P. 214–223. DOI: 10.1093/jole/lzad001. Advance access publication 18 April 2023.
12. Wang W., Mao L., Wang R., Min B.-C. 2023 — W. Wang, L. Mao, R. Wang, B.-C. Min. Multi-Robot Cooperative Socially-Aware Navigation Using Multi-Agent Reinforcement Learning // arXiv. Available at: <https://arxiv.org/abs/2309.15234> [Дата обращения: 03.04.2024].
13. Wen M., Kuba J. G., Lin R., Zhang W., Wen Y., Wang J., Yang Y., et al. 2022 — M. Wen, J. G. Kuba, R. Lin, W. Zhang, Y. Wen, J. Wang, Y. Yang, et al. Multi-Agent Reinforcement Learning is a Sequence Modeling Problem // arXiv. Available at:  
<https://arxiv.org/abs/2205.14953> [Дата обращения: 17.02.2024].
14. Zhuge M., Liu H., Faccio F., Ashley D. R., Csordás R., Gopalakrishnan A., Hamdi A., Hammoud H. A. K., Herrmann V., Irie K., Kirsch L., Li B., Li G., Liu S., Mai J., Piękos P., Ramesh A., Schlag I., Shi W., Stanić A., Wang W., Wang Y., Xu M., Fan D.-P., Ghanem B., Schmidhuber J. 2023 — M. Zhuge, H. Liu, F. Faccio, D. R. Ashley, R. Csordás, A. Gopalakrishnan, A. Hamdi, H. A. K. Hammoud, V. Herrmann, K. Irie, L. Kirsch, B. Li, G. Li, S. Liu, J. Mai, P. Piękos, A. Ramesh, I. Schlag, W. Shi, A. Stanić, W. Wang, Y. Wang, M. Xu, D.-P. Fan, B. Ghanem, J. Schmidhuber. Mindstorms in Natural Language-Based Societies of Mind // arXiv. Available at: <https://arxiv.org/abs/2305.17066> [Дата обращения: 05.12.2023].

## 6. Приложение

1. Introduction to OpenAI Platform. OpenAI. URL:  
<https://platform.openai.com/docs/introduction> [Дата обращения: 05.12.2023].

2. Research on AI Agents Communication. GitHub. URL:  
<https://github.com/TagirRamilevich/research-ai-agents-communication> [Дата обращения:  
25.05.2024].