

ПРИМЕНЕНИЕ АДАПТИВНОГО RL-АГЕНТА ДЛЯ ДИНАМИЧЕСКОГО РЕГУЛИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПОТОКОВ В ММО-СИМУЛЯЦИИ

THE USE OF AN ADAPTIVE RL AGENT FOR DYNAMIC REGULATION OF ECONOMIC FLOWS IN AN MMO SIMULATION

**M. Figurov
B. Mishchuk**

Summary. Regulation of economic flows in closed virtual economies is one of the most urgent tasks in the modern field of information technology. Unlike in the real world, the value of a virtual currency directly depends on the proper configuration of its sources of origin and distribution, called faucets and sinks. Mistakes in regulating economic flows lead to economic destabilization, a sharp spike in inflation or deflation, and the collapse of the virtual market, which in the long run leads to a decrease in the number of users and large financial losses. Traditionally, the balance of the virtual economy is handled by a whole department of specialists who respond promptly to shocks. This paper explores the possibility of automating this process using the Reinforcement Learning method.

This article discusses the development and application of an adaptive RL agent capable of managing key economic parameters in real time to stabilize the money supply in the virtual economy of the EvE online MMO game.

The software package developed because of the research demonstrates the ability not only to maintain system stability, but also to apply various strategies for emergency response to crises, which opens prospects for the development of self-regulation mechanisms in virtual economies.

Keywords: virtual economy, reinforcement learning, Bellman equation, mathematical modeling, dynamic regulation, neural networks.

Фигуров Максим Олегович

Аспирант,

Балтийский федеральный университет им. И. Канта

impromaximus@gmail.com

Мишчук Богдан Ростиславович

кандидат физико-математических наук, доцент,

Балтийский федеральный университет им. И. Канта

b.mishchuk@yandex.ru

Аннотация. Регулирование экономических потоков в замкнутых виртуальных экономиках является одной из наиболее актуальных задач в современной сфере информационных технологий. В отличие от реального мира, ценность виртуальной валюты прямо зависит от грамотной настройки источников её возникновения и сбыта, именуемых faucets и sinks. Ошибки в регулировании экономических потоков приводят к дестабилизации экономики, резкому скачку инфляции либо дефляции и обвалу виртуального рынка, что в долгосрочной перспективе приводит к снижению числа пользователей и большим финансовым потерям. Традиционно балансом виртуальной экономики занимается целый отдел специалистов, которые оперативно реагируют на шоковые воздействия. В настоящей работе исследуется возможность автоматизации данного процесса с помощью метода обучения с подкреплением (Reinforcement Learning).

В настоящей статье рассмотрен вопрос разработки и применения адаптивного RL-агента, способного в режиме реального времени регулировать ключевые экономические параметры для стабилизации денежной массы в виртуальной экономике ММО-игры EvE online.

Разработанный в результате исследования программный комплекс демонстрирует способность не только поддерживать стабильность системы, но и применять различные стратегии для экстренного реагирования на кризисы, что открывает перспективы для разработки механизмов саморегуляции в виртуальных экономиках.

Ключевые слова: виртуальная экономика, reinforcement learning, уравнение Беллмана, математическое моделирование, динамическое регулирование, нейросети.

Введение

Виртуальная экономика в онлайн-играх и мета-сервисах является симулятором, прообразом которого служит экономика реального мира, за исключением того, что вся генерация валюты и ресурсов внутри системы полностью подконтрольна разработчикам и не привязана к чему-либо материальному [1]. Тем не менее доход с неё вполне реален: к примеру, согласно финансовым отчетам по состоянию на 2022 год, популярная ММО игра World of Warcraft принесла разработчикам порядка

12.8 миллиардов долларов [2]. Рынок игровой индустрии демонстрирует экспоненциальный рост за последнее десятилетие с лидирующим трендом на создание виртуальных миров и онлайн-сервисов с собственной экономикой [3]. Однако для достижения стабильности виртуальной экономики без ущерба для игрового процесса, разработчикам приходится балансировать замкнутую экономическую систему таким способом, чтобы она исключала критические состояния, такие как инфляция или дефляция, а также была максимально прозрачной и справедливой по отношению к своим участникам.

Объектом исследования в данной работе выступает процесс динамического регулирования экономических потоков в модели, построенной на основе открытых данных по экономике MMO-игры EVE Online. Особое внимание было уделено разработке и анализу адаптивного RL-агента, способного автономно управлять ключевыми экономическими параметрами. Результаты работы данного программного комплекса были представлены в научной работе, где они служат примером эффективного применения машинного обучения с подкреплением при решении задачи динамического регулирования потоков денежных масс внутри замкнутой виртуальной экономики.

Целью настоящей научно-исследовательской работы является проектирование и разработка программного комплекса на базе RL для решения проблемы для динамического регулирования экономических потоков в MMO-симуляции.

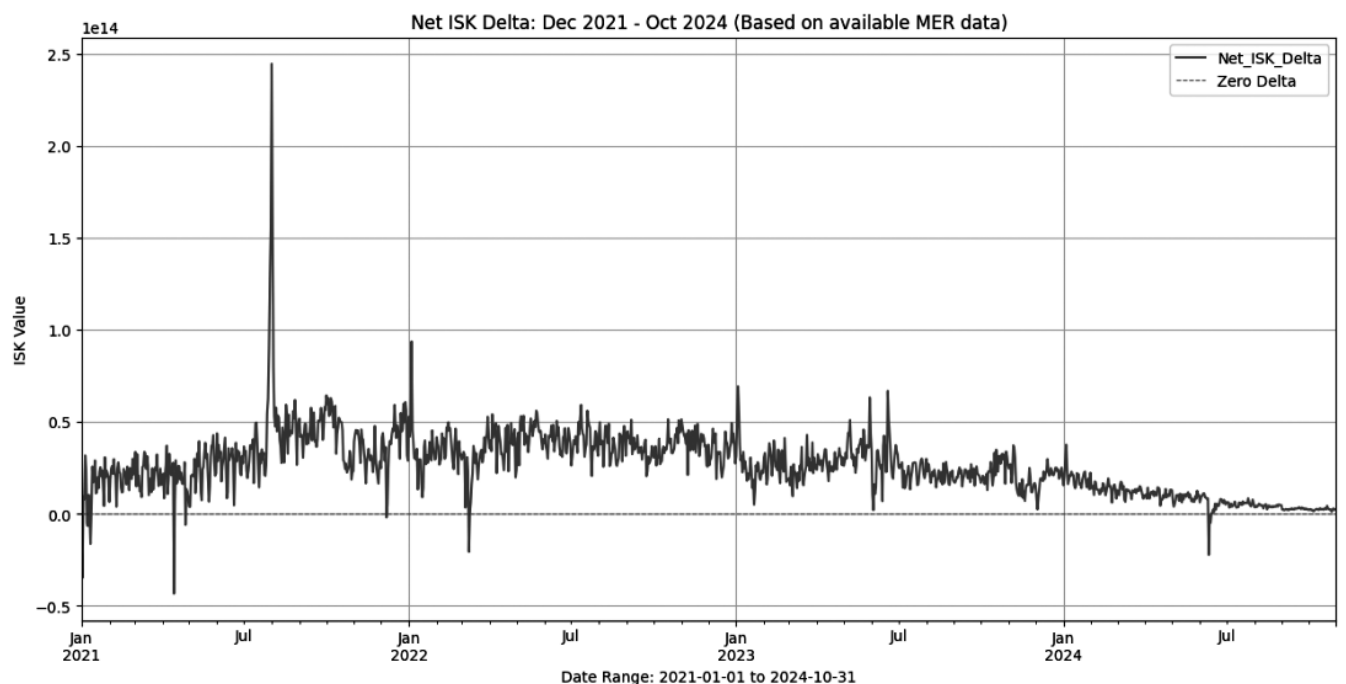
Задачами исследования являются:

1. Анализ экономической системы EVE Online и выявление основных проблем с её регулированием.
2. Разработка математической модели виртуальной экономической среды для RL-агента и его обучение.
3. Разработка прототипа программного комплекса динамического регулирования экономических потоков в виртуальной экономике и анализ результатов его работы.

Методы исследования

Виртуальная экономика многопользовательских онлайн-проектов обычно устроена таким образом, что внутриигровые механики проекта направлены на получение или трату игроками виртуальной валюты. Источники получения валюты и точки её сбыта строго контролируются разработчиками и именуются «кранами» (англ. faucets) и «раковинами» (англ. sinks) [4]. Ключевая задача разработчиков — настроить экономические потоки таким образом, чтобы не допускать возникновения неконтролируемой инфляции или дефляции валюты, особенно при возникновении шоковых воздействий на рынок, таких как глобальные обновления проекта, эксплойты системы или масштабные внутриигровые события.

Показательным примером сложной, управляемой игроками экономики является MMO EVE Online. Для анализа её актуального состояния разработчики ежемесячно публикуют отчеты MER (Monthly Economic Reports), представляющие собой открытые данные по различным аспектам актуальной внутриигровой экономики [5]. Несмотря на продуманность ручного баланса экономики, он не всегда даёт оптимальный результат. Для подтверждения этого тезиса рассмотрим агрегированные данные из отчетов MER. В качестве количественного показателя экономической стабильности возьмем чистый денежный поток (Net ISK Delta, где ISK — название внутриигровой валюты). Он отражает разницу между всеми «кранами» и «раковинами» и напрямую демонстрирует



Обработанная таблица НОВОГО ФОРМАТА сохранена в: C:\EVE_MER_Project_Data\processed_daily_isk_flows_pivot_NEW_FORMAT.csv

Рис. 1. Динамика показателя Net ISK Delta в EVE Online с 2021 по 2024 год. Несмотря на приближение к нулевому балансу, наблюдаются пиковые значения

тенденции роста или сокращения общей денежной массы в игре. На основе данных MER был построен временной ряд показателя Net ISK Delta в период с 2021 по 2024 год (см. рис. 1).

Net ISK Delta в основном держится около нуля или немного выше, что является более устойчивым состоянием, чем доминанция выраженных инфляционных и дефляционных трендов. Однако экономика остаётся неустойчивой, о чем свидетельствует наличие множественных экстремальных пиковых значений на графике, когда денежная масса стремительно увеличивается или сокращается под воздействием внутриигровых событий. Например, отрицательное пиковое значение в апреле 2022 года совпадает по времени с последствиями глобального военного конфликта («World War Bee 2») [6]. В июле 2021 года разработчики провели заключительный этап ликвидации устаревших космических станций игроков и заменили их на новую систему, при этом проведя программу компенсации. В результате игроки стали массово продавать компенсационные предметы, что привело к вливанию триллионов ISK в экономику EVE. Этот случай — пример экзогенного экономического шока, который также зафиксирован на официальном графике MER, представленном разработчиками (см. рис. 2).

Резюмируя проведенный анализ, даже в относительно стабильный период, экономика подвержена резким шокам, на которые ручное управление не всегда способ-

но среагировать. Предполагаемая RL-система призвана решить эту проблему путем динамической адаптации экономических параметров, обеспечивая немедленную реакцию на любые предпосылки дестабилизации экономических потоков внутри игры.

Динамическое регулирование экономики представляет собой Марковский процесс принятия решений, определяемый кортежем S, A, P, R, γ , где S — пространство состояний, A — пространство действий, P — функция вероятностей перехода, R — функция вознаграждения, γ — коэффициент дисконтирования [7]. В рассматриваемой экономической среде:

- Состояние определяется как $s_t \in S$ и включает в себя всю историю показателя Net ISK Delta (d_t, d_{t-1}).
- Действие определяется как $a_t \in A$ и определяет величины изменений экономических параметров.
- Функция перехода определяется как $P(s_{t+1} | s_t, a_t)$ и описывает динамику системы, учитывая стохастичность среды. Net ISK Delta моделируется как стохастический авторегрессионный процесс:

$$d_{t+1} = \alpha \cdot d_t - \beta \cdot f(a_t) + \varepsilon_t \quad (1)$$

где α — коэффициент инерции, β — вектор эффективности, а ε_t — случайный шум ($\varepsilon_t \sim N(0, \sigma^2)$).

- Функция вознаграждения определяется как R_t и призвана стабилизировать экономику, обеспечивая плавность регулирования:

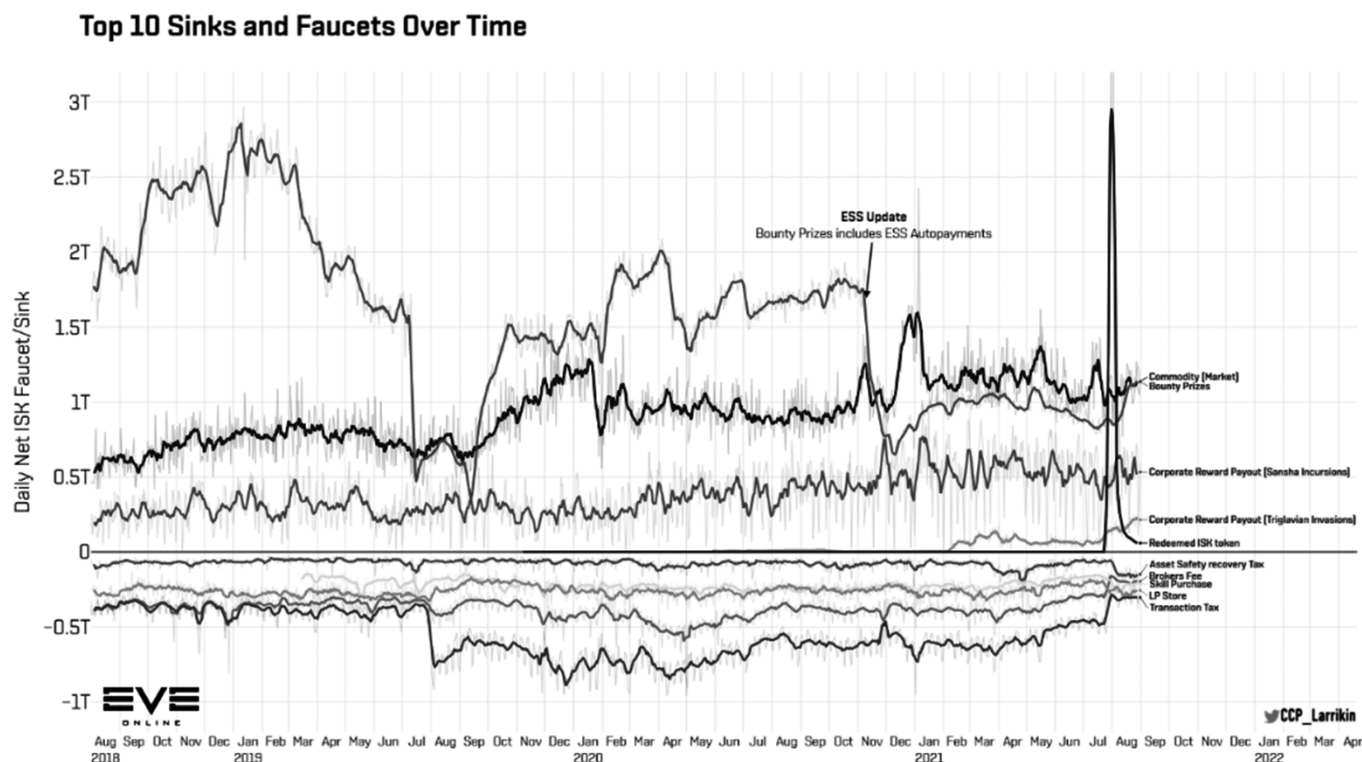


Рис. 2. График динамики денежных масс. Черный пик вверх на графике отражает резкое вливание новой валюты в экономику после обновления в июле 2021 года

$$R_t = w_d \cdot \exp(-k_d \cdot |d_{t+1}|) - w_a \cdot \|a_t - a_{t-1}\|^2 \tag{2}$$

Цель RL-агента — найти оптимальную политику $\pi^*(a|s)$, максимизирующую ожидаемое дисконтированное вознаграждение $G_t = \sum \gamma^k \cdot R_{t+k+1}$. Фундаментом для нахождения такой политики служит уравнение оптимальности Беллмана:

$$Q^*(s, a) = E_{s' \sim p} [R(s, a, s') + \gamma \cdot \max_{a'} Q^*(s', a')] \tag{3}$$

Из-за непрерывности пространств состояний и действий аналитическое решение уравнения оптимальности Беллмана в данном случае не представляется возможным, однако нахождение оптимальной политики возможно через функциональную аппроксимацию с помощью нейронных сетей. В данном случае агент взаимодействует со средой, описанной уравнениями (1) и (2) и итеративно обновляет веса сетей, чтобы через обучение нейросетевого агента найти численное приближенное решение уравнения (3).

Для реализации задачи разработки RL-системы были использованы современные фреймворки [8] (см. табл. 1). Проект был реализован на языке Python 3.9.

Таблица 1.

Технологический стек проекта

Компонент	Инструмент/Библиотека	Назначение в проекте
Среда моделирования	Gymnasium	Создание среды, стандартизация API
RL-агент	Stable-Baselines3 (PPO)	Реализация, обучение и оценка RL-агента
Deep Learning	PyTorch	Нейросетевой бэкэнд для Stable-Baselines3
Работа с данными	Pandas, NumPy	Обработка временных рядов
Визуализация	Matplotlib, Seaborn	Построение графиков для представления результатов

Стратегия агента (ppo_eve_econ_5000000.zip)

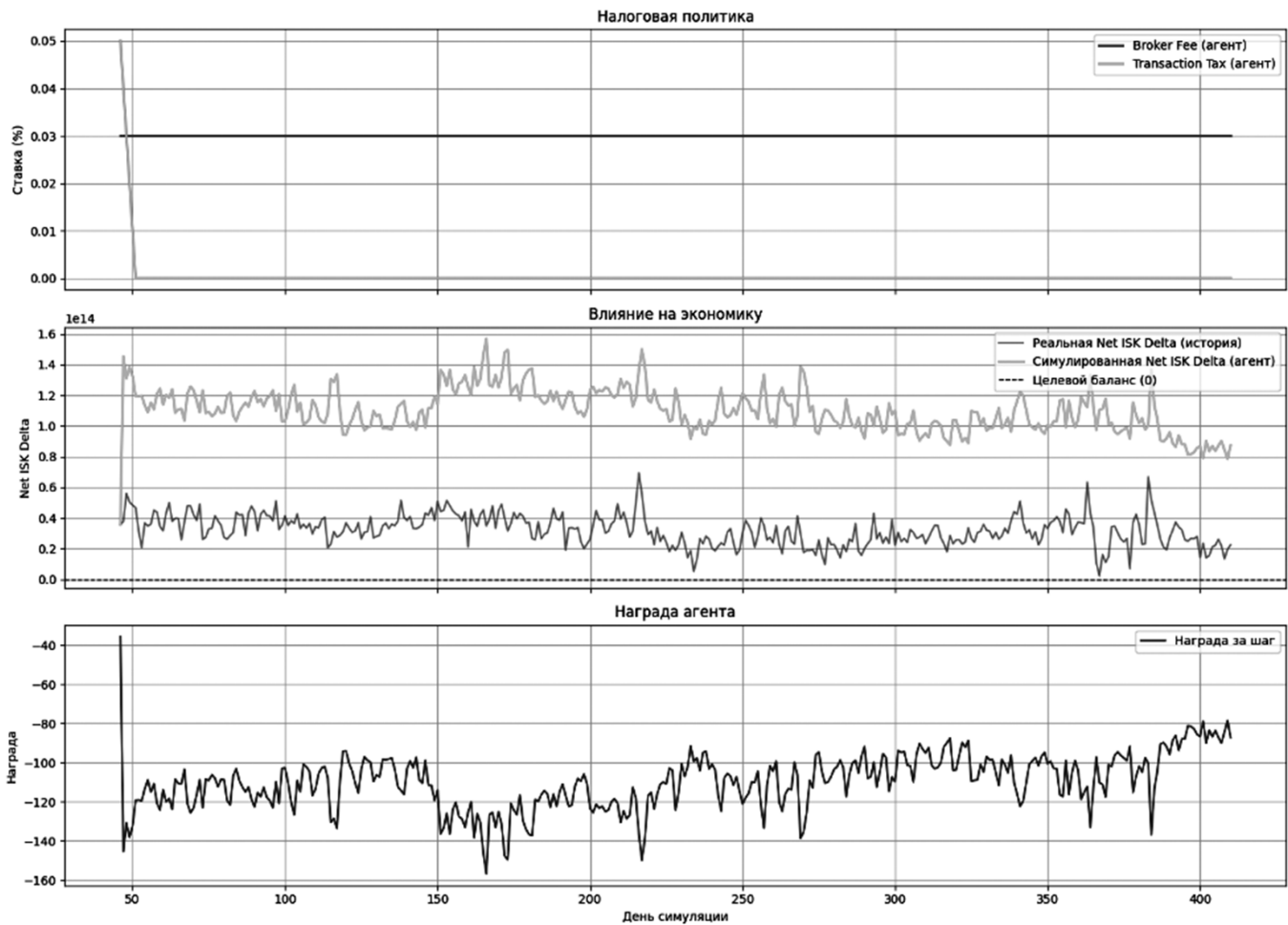


Рис. 3. Первая итерация RL-агента (PPO). Темно-серая линия на среднем графике соответствует реальным данным, светло-серая — данным полученным в результате симуляции

В качестве алгоритма обучения в первой итерации RL-агента был выбран PPO [9], цель которого заключалась в регуляции налоговых ставок на транзакции для снижения инфляции, однако это привело к противоположному результату (см. рис. 3). Агент научился симулировать Net ISK Delta на основе реальных исторических данных, но постоянно выбирал снижение налоговой ставки как оптимальную стратегию, что способствовало дальнейшему усилению инфляционных процессов.

В процессе тестирования новых итераций агента были выявлены две основные проблемы данного этапа разработки — недостаточная сложность экономической модели и избыточно высокие штрафы за апробацию разных действий при обучении. Для достижения оптимального уровня налогообложения экономической движок среды был переписан с учетом закона Кривой Лаффера. Также алгоритм PPO был заменен на SAC (Soft Actor-Critic), ввиду его больше пригодности для обеспечения плавной настройки в непрерывном пространстве действий [10]. В результате была получена новая тестовая итерация, которая продемонстрировала успешное решение задачи снижения инфляции (см. рисунок 4).

Основной проблемой данной итерации является недостаточная сложность механизма регуляции — в EVE Online нет такого понятия как «единый налог», поскольку регуляция денежных масс осуществляется через большое количество вариаций налоговых ставок. С целью повышения практической значимости разрабатываемой системы были введены дифференцированные налоги на основе реальных данных из отчетов MER:

1. `market_tax_multiplier` — действие, регулирующее одновременно две основных налоговые ставки (Transaction Tax и Brokers Fee), «раковина» для денежных масс.
2. `bounty_payout_multiplier` — главный инструмент монетарной политики, контроль размера наград за выполнение заданий, «кран» для денежных масс.
3. `direct_intervention` — прямой контроль денежной массы при возникновении экстремальных шоков, диапазон [-1.0, 1.0]

Новая версия RL-агента успешно справилась с задачей динамического регулирования экономики (см. рис. 5)

Разработанный агент продемонстрировал способность стабильно удерживать Net ISK Delta около целевого

Стратегия агента (sac_eve_econ_final_500000.zip)

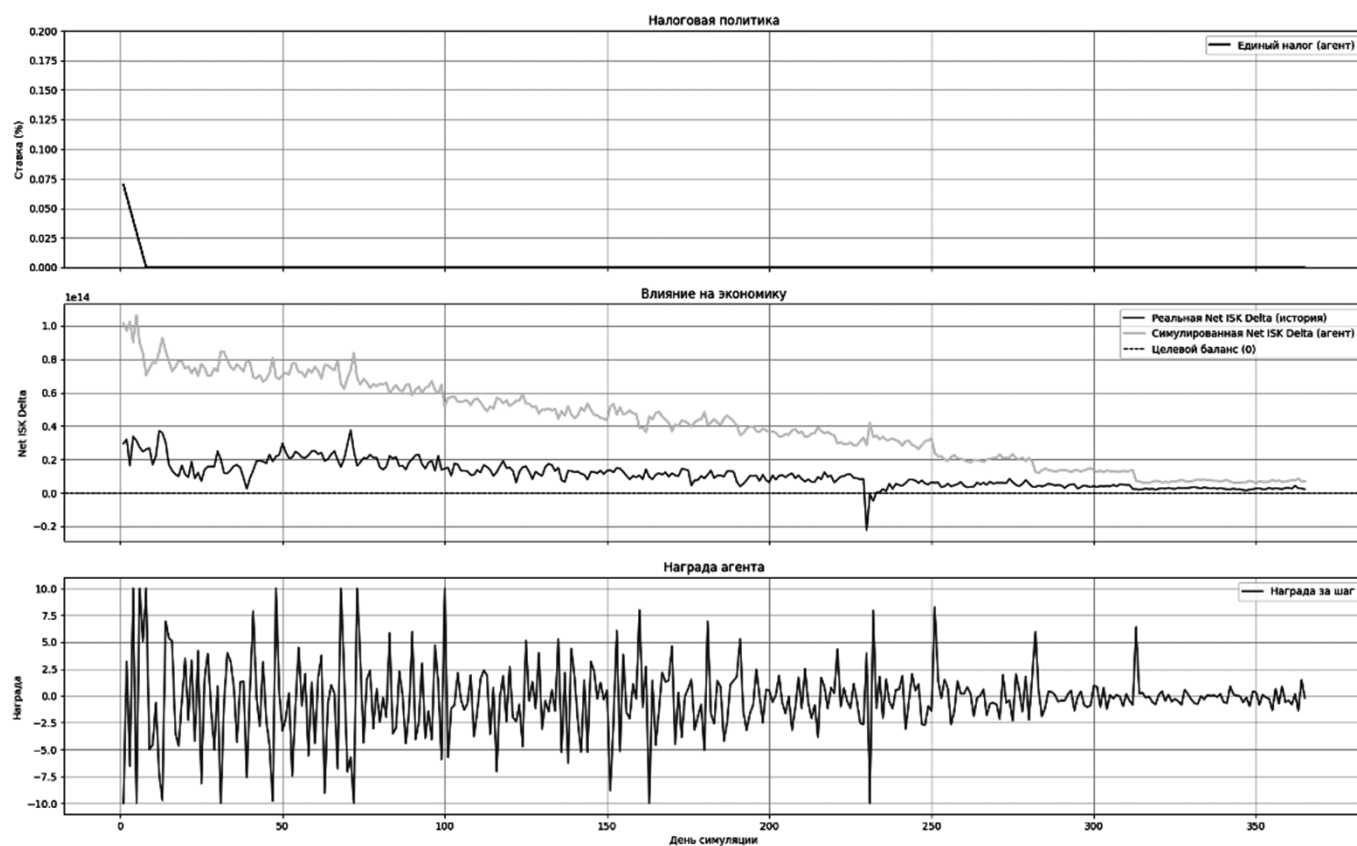


Рис. 4. Вторая итерация RL-агента (SAC). Темно-серая линия на среднем графике соответствует реальным данным, светло-серая — данным полученным в результате симуляции

Результаты Работы RL-Агента (v2) по Управлению Экономикой

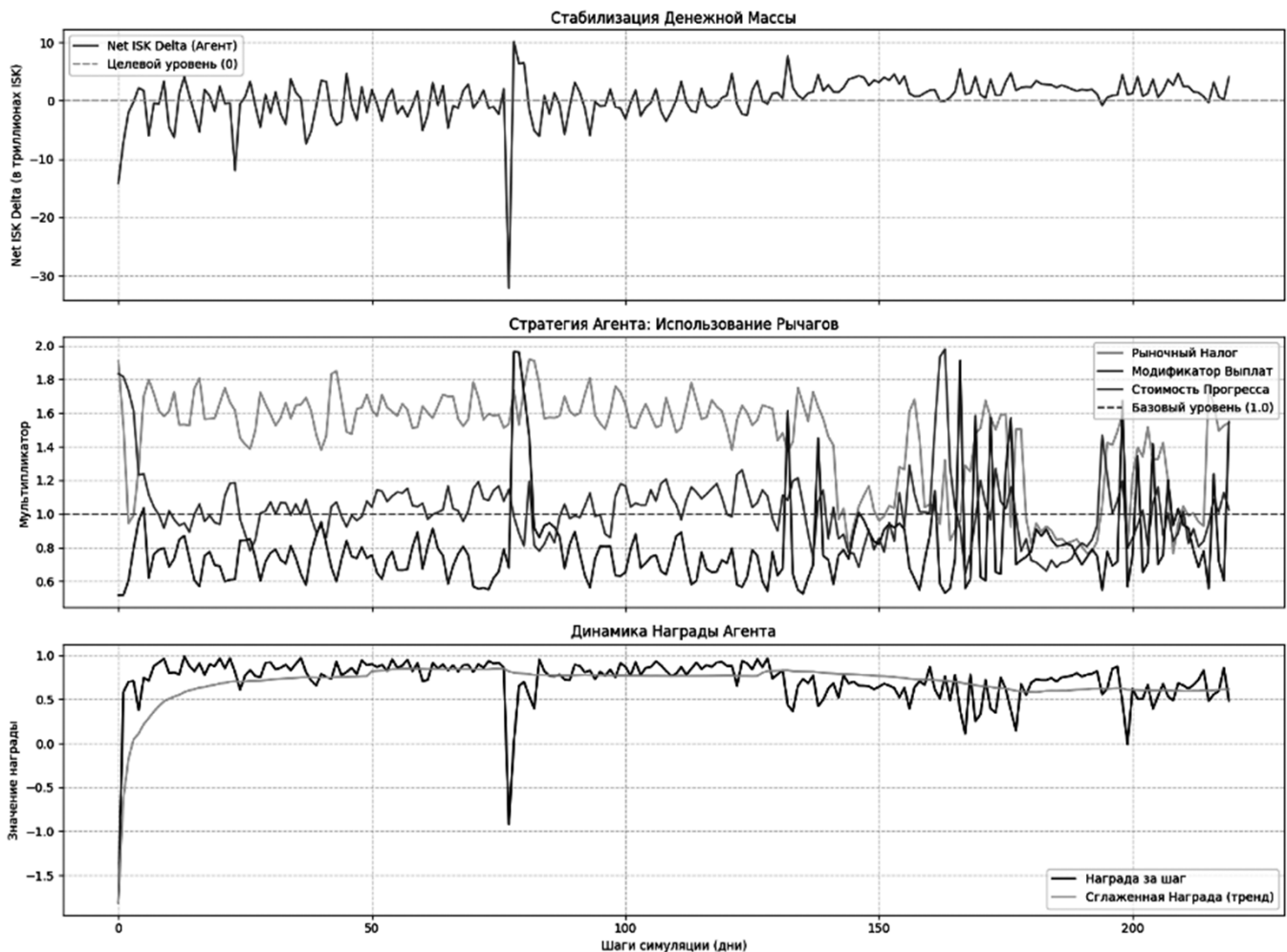


Рис. 5. Новая версия RL-агента. Темно-серая линия на верхнем графике отражает изменение Net ISK Delta

значения, а также оперативно реагировать на кризисы — на графике выше видно, как агент успешно скорректировал дефляционный скачок в краткий срок. В процессе обучения RL-агент принял решение установить на постоянной основе высокие налоги при значительном сокращении выплат игрокам, что в долгосрочной перспективе игрового процесса, вероятно, приведет к сокращению пользовательской базы. Чтобы исправить данную проблему, в функцию награды при обучении модели был добавлен новый компонент — штраф за существенное отклонение параметров регулирования от базовых значений. Для достижения стабильной и прозрачной экономики был также введен штраф за резкие изменения значений при принятии решений RL-агентом (см. рис. 6).

Обновленная версия RL-агента успешно поддерживает объем денежных масс около целевого значения, эффективно справляется с шоками и обеспечивает высокий уровень прозрачности для игроков. Налоговые ставки стали более предсказуемы и всегда поддерживаются

на стабильном уровне, что позволяет игрокам строить свои бизнес-модели вокруг данного факта и минимизирует для них риски, связанные с резкими изменениями. Расходы на игровую прогрессию поддерживаются на одном и том же уровне, но могут быть резко увеличены при реакции на кризисную ситуацию, выступая как инструмент экстренного реагирования. Регулирование наград используется для тонкой настройки, компенсируя мелкие отклонения для поддержания основной политики агента. В итоге финальная версия RL-агента успешно разделяет инструменты настройки экономики и экстренно реагирует на кризисы, динамически выполняет логичные и предсказуемые действия по успешной стабилизации объемов денежных масс, а также создает интересный и живой игровой процесс для всех участников виртуальной экономики внутри игры.

Результаты исследований

В результате исследований были получены следующие результаты:

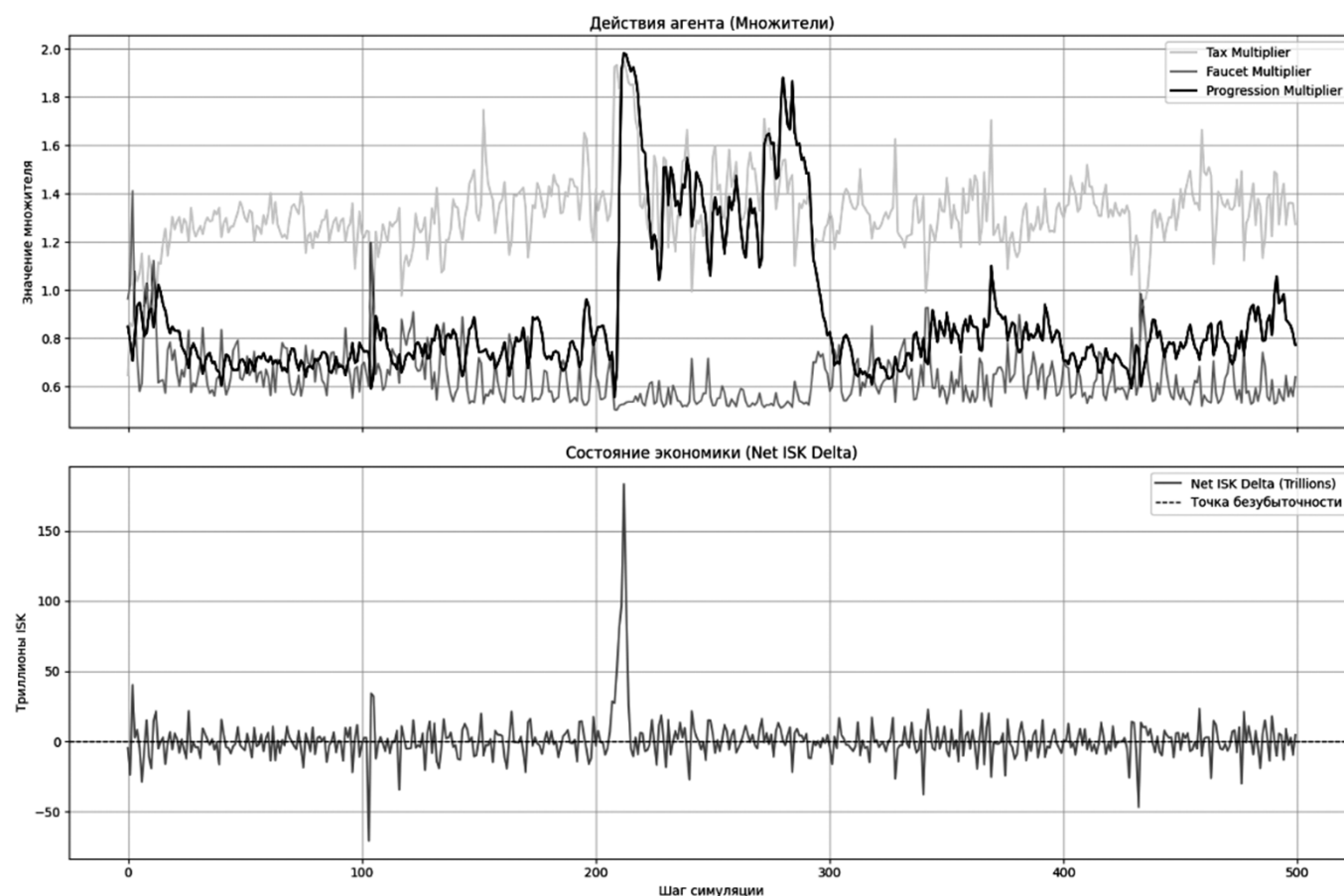


Рис. 6. Финальная итерация RL-агента. Верхний график отражает действия агента в разные промежутки времени

- Проведен анализ экономической системы EVE Online, выявлены основные проблемы с её регулированием.
- Разработана математическая модель виртуальной экономической среды, послужившая основой для обучения RL-агента, предназначенного для регулирования экономики.
- Создан прототип программного комплекса для динамического регулирования экономических потоков в виртуальной экономике и продемонстрирована эффективность его использования.

Заключение

В рамках научно-исследовательской работы был разработан прототип программного комплекса динамического регулирования экономических потоков в виртуальной экономике. В ходе работы проведен анализ экономической системы EVE Online, разработана математическая модель виртуальной экономической среды, на основе которой был обучен RL-агент по регулированию экономики.

ЛИТЕРАТУРА

1. Lehdonvirta V., Castronova E. Virtual Economies: Design and Analysis. — Cambridge: The MIT Press, 2014. — P. 9–12
2. «Легенда, рожденная из хаоса» [Электронный ресурс]. URL: <https://www.kommersant.ru/doc/7325460/> (дата обращения: 22.06.2025).
3. Castronova E. Exodus to the virtual world. — St. Martin's Press, 2007
4. «Sinks & Faucets: Lessons on Designing Effective Virtual Game Economies» [Электронный ресурс]. URL: <https://medium.com/1kxnetwork/sinks-faucets-lessons-on-designing-effective-virtual-game-economies-c8daf6b88d05/> (дата обращения: 25.06.2025)
5. CCP Games. Monthly Economic Reports [Электронный ресурс]. URL: <https://www.eveonline.com/news/t/monthly-economic-reports/> (дата обращения: 28.06.2025)
6. «World War Bee 2 A beginner's guide to EVE Online's epic inter-player war» [Электронный ресурс]. URL: <https://tagh.wordpress.com/wp-content/uploads/2022/02/world-war-bee-2-brief-final.pdf/> (дата обращения: 01.07.2025)
7. Sutton R.S., Barto A.G. Reinforcement learning: An introduction. MIT press, 2015. — P.53–80
8. Raffin A. et al. Stable-baselines3: Reliable reinforcement learning implementations // The Journal of Machine Learning Research. — 2021. — T.22. — №. 1. — С. 1–8.
9. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal policy optimization algorithms // arXiv preprint arXiv:1707.06347. — 2017.
10. Уиндер Ф. Обучение с подкреплением для реальных задач. — СПб.: БХВ-Петербург, 2023. — 400 с.

© Фигуров Максим Олегович (impromaximus@gmail.com); Мишук Богдан Ростиславович (b.mishchuk@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»