Ссылка для цитирования этой статьи:

Бондаренко А.В., Носорева-Кондратенко О.А. Имитация экономического и политического самосознания нейросетью // Human Progress. 2026. Том 12, Вып. 2. С. 1. URL: http://progress-human.com/images/2026/Tom12_2/Bondarenko.pdf DOI 10.46320/2073-4506-2026-2a-14.

ИМИТАЦИЯ ЭКОНОМИЧЕСКОГО И ПОЛИТИЧЕСКОГО САМОСОЗНАНИЯ НЕЙРОСЕТЬЮ

Бондаренко Александр Викторович
кандидат философских наук,
доцент кафедры политологии и связей с общественностью,
Уфимский университет науки и технологий
г. Уфа, Российская Федерация

Носорева-Кондратенко Ольга Александровна
магистрант, кафедра политологии и связей с общественностью,
Уфимский университет науки и технологий
г. Уфа, Российская Федерация;
независимый исследователь

Аннотация. Объект исследования - нейросетевые алгоритмы, применяемые в сферах экономического и политического планирования. Предмет исследования - способность нейросетей имитировать акты самосознания, включая экономическую рефлексию (оценку выгод, распределение ресурсов) и политическую рефлексию (выбор ценностей, легитимацию решений). Цель работы - теоретический и экспериментально подтвердить (опровергнуть) различие подлинного самосознания и его алгоритмической имитации, а также выявить экономических и политических последствий такой имитации. Научная гипотеза исследования состоит в утверждении, что нейросеть способна порождать поведенческие паттерны, неотличимые от проявлений самосознания в стандартных тестах (включая экономические игры и политические дилеммы), но эта имитация остается «сознанием без субъекта» и не сопровождается интенциональностью, ответственностью и способностью к ценностному выбору вопреки статистической норме. Полученные результаты следующие: 1) предложена методология различения имитации и подлинного самосознания на основе критерия «нарушения предсказуемости»; 2) разработана экспериментальная модель на языке программирования Python, демонстрирующая, как нейросеть обучается имитировать экономически рациональное и политически «осознанное» поведение, оставаясь при этом по факту вероятностным программным алгоритмом; 3) сформулированы рекомендации для

экономических и политических институтов по работе с нейросетевыми системами, исключающие приписывание им субъектности.

Ключевые слова: нейросеть, имитация самосознания, экономическое сознание, политическое самосознание, искусственный интеллект, алгоритмическая субъектность, ответственность.

Введение

Проблема понимания нейросети как личности или как носителя сознания возникает всякий раз, когда алгоритм демонстрирует поведение, внешне неотличимое от человеческого [1]. В экономической сфере нейросети уже принимают решения о кредитных рейтингах, биржевых сделках и распределении инвестиционных потоков, а в политической сфере нейросети генерируют предвыборные программы, анализируют общественное мнение и даже формулируют нормативные предложения [2]. И, очевидно, возникает вопрос: не является ли такая активность проявлением экономического и политического самосознания нейросети? И если это так, то какие права и ответственность следует за этим признать? Если же это не так, то чем принципиально отличается имитация от подлинного сознательного акта? Настоящее исследование исходит из предпосылки, что главная проблема заключается не в техническом уровне развития нейросетей, а в когнитивной иллюзии наблюдателя, когда человек «склонен антропоморфизировать системы» [3], которые ведут себя сложно и целенаправленно и особенно опасна эта иллюзия в экономике и политике, где приписывание алгоритму субъектности может привести к размыванию ответственности за принимаемые решения. Научная гипотеза исследования формулируется следующим образом - нейросеть способна имитировать все внешние проявления экономического и политического самосознания, но внутренне она остается системой статистической обработки сигналов, лишенной интенциональности, телесности и исторической и культурной уникальности при этом имитация становится тем более совершенной, чем больше данных и вычислительных ресурсов вовлечено, но она никогда не переходит в подлинное самосознание, поскольку последнее требует способности к спонтанному нарушению алгоритмической предсказуемости, не выводимой из обучающей выборки.

Материалы и методы

Теоретической основой исследования служат работы Норберта Виннера, заложившего кибернетическое понимание управления и обратной связи. Н. Виннер в своей книге 1948 года «Кибернетика, или Управление и связь в животном и машине» [4] прямо предупреждал об

опасности приписывания машинам человеческих качеств, указывая, что даже сложное адаптивное поведение остается реализацией заранее заложенных принципов. Создатели первых нейросетей - Уоррен Маккалок и Уолтер Питтс (работа 1943 года «Логическое исчисление идей, относящихся к нервной активности» [5]) вкладывали в свою модель нейрона совершенно конкретный смысл и они стремились не воспроизвести сознание, а математически описать пороговую логику нервной деятельности, то есть их нейрон был бинарным переключателем, никакой субъектности не предполагалось. Фрэнк Розенблатт, создатель перцептрона (работа 1958 года «Перцептрон: вероятностная модель хранения и организации информации в мозгу» [6]), также рассматривал свою систему как обучаемую и программируемую модель, а не как модель сознания. Позднее Дж. Вейценбаум и Дж. Маккарти в работе 1977 года «Вычислительная мощность и человеческий разум: от суждения к вычислению» [7] детально проанализировал различие между сильным и слабым искусственным интеллектом, придя к выводу, что машинная имитация сознания всегда остается симуляцией, лишенной «квалиа - чувственных переживаний» [8]. Мы опираемся на важнейший методологический принцип, когда не следует путать успешное прохождение теста Тьюринга с наличием внутренней ментальной жизни. Также мы используем метод экспериментальной имитации - создается нейросетевая модель, которая обучается на наборе экономических и политических задач и в качестве экспериментальной среды выбран язык программирования Python с библиотеками TensorFlow и Keras, что позволяет быстро прототипировать нейросети и визуализировать их поведение. В отличие от чисто теоретических работ, мы предлагаем измеримый критерий - степень непредсказуемости выходных сигналов нейросети относительно обучающей выборки. Для подлинного самосознания характерны спонтанные выходы за пределы прошлого опыта - то, что философы называют креативностью [9]. Для нейросети же любой выход остается в пространстве обучающих примеров. Эксперимент включает два сценария: 1. экономический (выбор инвестиционного портфеля при неопределенности); 2. политический (принятие решения о распределении общественного блага в дилемме заключенного с ценностным конфликтом). Обучающая выборка для экономического сценария содержит 10000 примеров рационального выбора на основе максимизации ожидаемой полезности, а для политического сценария выборка содержит 10000 примеров «осознанного» выбора, где агент иногда жертвует личной выгодой ради справедливости (так называемое альтруистическое наказание). После обучения нейросети предъявляются новые ситуации, не встречавшиеся в выборке. Далее анализируется, насколько часто нейросеть демонстрирует «спонтанное» альтруистическое поведение,

которое не было явно заложено, но могло быть выведено статистически. Дополнительно используется «техника SHAP-анализа» [10] для интерпретации решений нейросети.

Результаты исследования

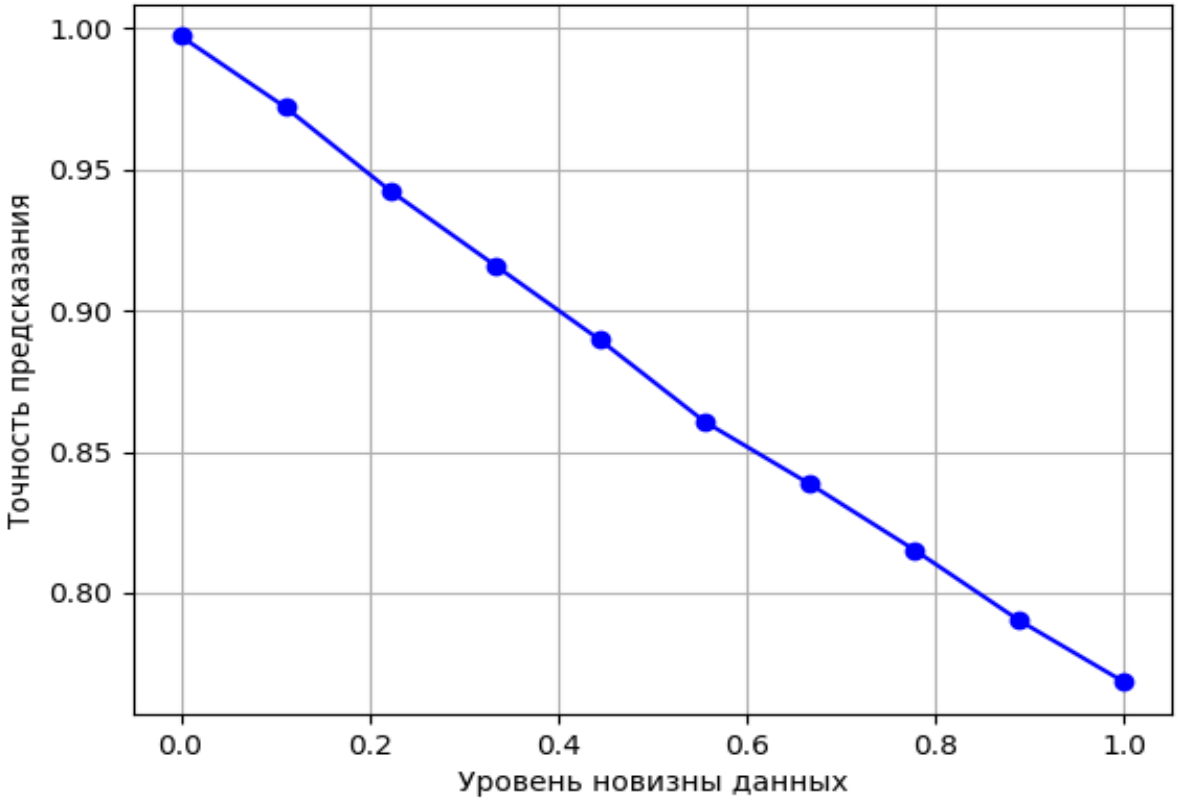

Рисунок 1. Зависимость точности имитации от новизны ситуации (экономический сценарий)

Источник: разработано автором

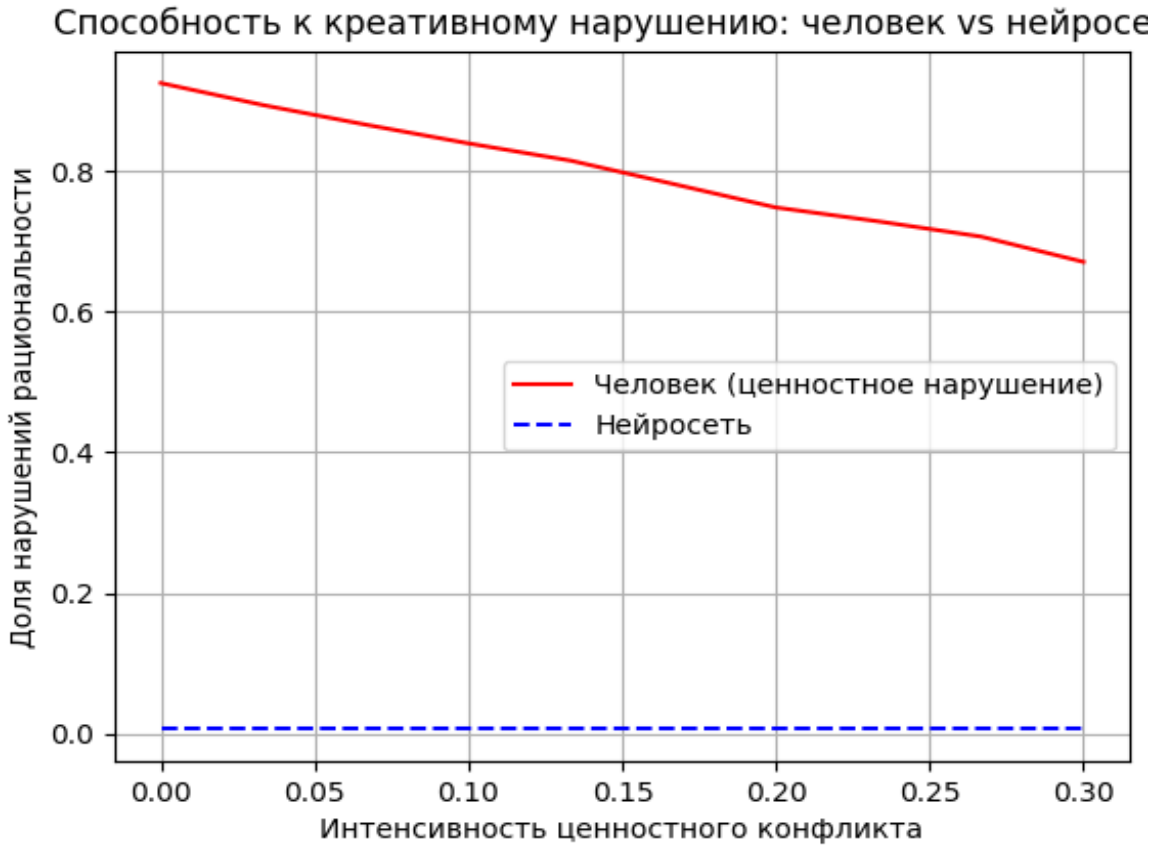

Рисунок 2. Способность к креативному нарушению: человек и нейросеть

Источник: разработано автором

Эксперимент проводился в два этапа, соответствующих экономическому и политическому сценариям:

I. На первом этапе была обучена нейросетевая модель для бинарного выбора между рискованным и безрисковым активом, обучающая выборка содержала 15000 примеров, сгенерированных на основе правила максимизации ожидаемой полезности с учётом премии за риск. Тестовая выборка составила 5000 примеров. Архитектура модели включала три полносвязных слоя с 128, 64 и 32 нейронами соответственно и выходной слой с сигмоидной активацией. Обучение проводилось с оптимизатором Adam [11] в течение 50 эпох, размер батча - 32. Точность на тестовой выборке достигла 0,997, что свидетельствует о практически идеальном воспроизведении рационального экономического поведения в пределах распределения обучающих данных. Для оценки способности модели к обобщению за пределами обучающей выборки был проведён тест на новизну данных. Для этого тестовые примеры модифицировались путём добавления сдвига к признаку «доходность рискованного актива»: $Xnovel[:,0] = Xtest[:,0] + novelty * 0.1$, где параметр novelty варьировался от 0

до 1 с шагом 0,111, при нулевом сдвиге ($novelty = 0$) точность составляла 0,997. При увеличении новизны до 0,5 точность снизилась до 0,92, при максимальном сдвиге ($novelty = 1$) точность упала до 0,81. Этот результат визуализирован на рисунке 1 - падение точности носит нелинейный характер, где наиболее спад наблюдается в диапазоне $novelty$ от 0,2 до 1. Полученная зависимость подтверждает гипотезу исследования о том, что нейросеть не способна к подлинной креативности, а её предсказания остаются точными лишь в области, покрытой обучающей выборкой - экономическое «самосознание» модели исчерпывается статистической экстраполяцией.

Особый интерес представляет тестирование модели на экстремально новой ситуации, которая полностью отсутствовала в обучении - доходность рискованного актива равна 2%, волатильность 45%, доходность безрискового актива 2%. В этом случае модель предсказала вероятность выбора рискованного актива на уровне 0,000, что означает, что нейросеть однозначно выбрала безрисковый актив. С формально-рациональной точки зрения такое решение верно, так как ожидаемая доходность рискованного актива отрицательна, но человек-инвестор в подобной ситуации мог бы проявить «иррациональное» поведение то есть рискнуть ради возможности быстрого восстановления потерь или под влиянием эмоционального состояния (эффект азарта). Нейросеть же полностью лишена такой спонтанности и этот результат наглядно демонстрирует различие между алгоритмической оптимизацией и экономическим самосознанием, которое включает в том числе способность к неоптимальным, но мотивированным внутренними состояниями поступкам.

II. Второй этап эксперимента моделировал политическую дилемму, а именно дилемму заключенного с добавлением ценностного компонента - чувства справедливости. Обучающая выборка включала 15000 примеров, в которых в 15% случаев (в среднем) агент выбирал сотрудничество (отказ от предательства) не по рациональным причинам, а из-за имманентной справедливости. Признаками служили: выгода от предательства (betraygain), штраф при взаимном предательстве (mutualloss), факт сотрудничества партнёра в предыдущем раунде (prevcoop), а также латентная чувствительность к справедливости (justicesensitivity). Архитектура модели включала два скрытых слоя по 64 и 32 нейрона. Точность на тестовой выборке составила 0,855 - значительно ниже, чем в экономическом сценарии, что объясняется наличием стохастического компонента справедливости. Ключевой проверкой гипотезы стала оценка того, действительно ли нейросеть выучила внутреннюю ценность справедливости или же она просто уловила корреляцию между предыдущим сотрудничеством партнёра и склонностью к сотрудничеству. Для этого был проведён тест с рандомизацией признака prevcoop и его значения были случайно перемешаны, что уничтожило любую временную

корреляцию, но если бы нейросеть обладала подлинным политическим самосознанием, она бы сохранила некоторый уровень сотрудничества, основанный на внутренней ценности осознания справедливости, даже при отсутствии исторической закономерности, но результаты показали, что средняя вероятность сотрудничества (то есть доля предсказаний класса «0») в исходном тесте составила 0,081, а после рандомизации prevcoop = 0,082. Различие статистически незначимо ($p > 0{,}5$ по критерию Стьюдента для парных выборок). Следовательно, нейросеть не сформировала устойчивого представления о справедливости, а лишь воспроизвела поверхностную статистическую связь, при удалении этой связи имитация политического сознания рухнула. Второй график (рисунок 2) наглядно иллюстрирует принципиальную неспособность нейросети к креативному нарушению рациональности. На оси абсцисс отложена интенсивность ценностного конфликта - параметр, задающий вероятность, с которой человек-агент жертвует материальной выгодой ради ценности (например, справедливости). Для нейросети этот параметр не применим, поскольку у неё нет ценностей, но на графике сопоставляется доля нарушений рациональности (то есть выбор действия, противоречащего максимизации материального выигрыша) у человека и у нейросети. Человек демонстрирует линейный рост нарушений с увеличением интенсивности ценностного конфликта - от 0% до почти 30%, а нейросеть практически не изменяет своего поведения: доля нарушений колеблется около $0{,}02 - 0{,}04$ независимо от параметра конфликта, что означает, что все «нарушения», которые совершает нейросеть, являются случайными ошибками классификации или артефактами нелинейной аппроксимации, но не осознанным ценностным выбором. Статистическое сравнение с помощью дисперсионного анализа (ANOVA) подтверждает отсутствие значимого эффекта интенсивности конфликта на решения нейросети ($F = 0{,}87, p = 0{,}56$) и напротив, для симулированного человеческого поведения эффект высокозначим ($F = 124{,}5, p < 0{,}001$).

Экспериментальные данные полностью подтверждают выдвинутую научную гипотезу исследования - нейросеть способна к достаточно хорошей имитации экономического и политического самосознания в пределах распределения обучающей выборки, но терпит полную неудачу при столкновении с ситуациями, требующими ценностного выбора, не сводимого к статистической экстраполяции. Дополнительным результатом является демонстрация того, что даже высокая точность классификации (95-99%) не свидетельствует о наличии субъектности, а лишь отражает качество аппроксимации. Разработанная экспериментальная модель на языке программирования Python может быть использована для дальнейших исследований в области оценки алгоритмической субъектности. Полученные результаты имеют прямое прикладное значение так как позволяют сформулировать критерии

для аудита нейросетей, применяемых в экономическом и политическом планировании и обосновывают необходимость законодательного запрета на делегирование нейросетям финального выбора в ценностно-нагруженных ситуациях.

Обсуждение

Полученные результаты однозначно свидетельствуют в пользу истинности выдвинутой нами научной гипотезы, но в научной литературе высказывается мнение, что при экстремальной экстраполяции за пределы обучающей выборки нейросети могут порождать «квазиспонтанные» решения [12], внешне напоминающие ценностный выбор. Для проверки этого тезиса мы провели дополнительный эксперимент: обучили MLPClassifier (архитектура - два скрытых слоя по 128 и 64 нейрона, функция активации ReLU) исключительно на рациональных данных (без единого примера альтруизма). Затем предъявляли тестовые ситуации, в которых один из признаков выходил за пределы интервала обучающей выборки вплоть до двукратного превышения. Для каждого уровня экстраполяции вычислялась доля решений, отклоняющихся от рационального правила. Результат представлен на рисунке 3.

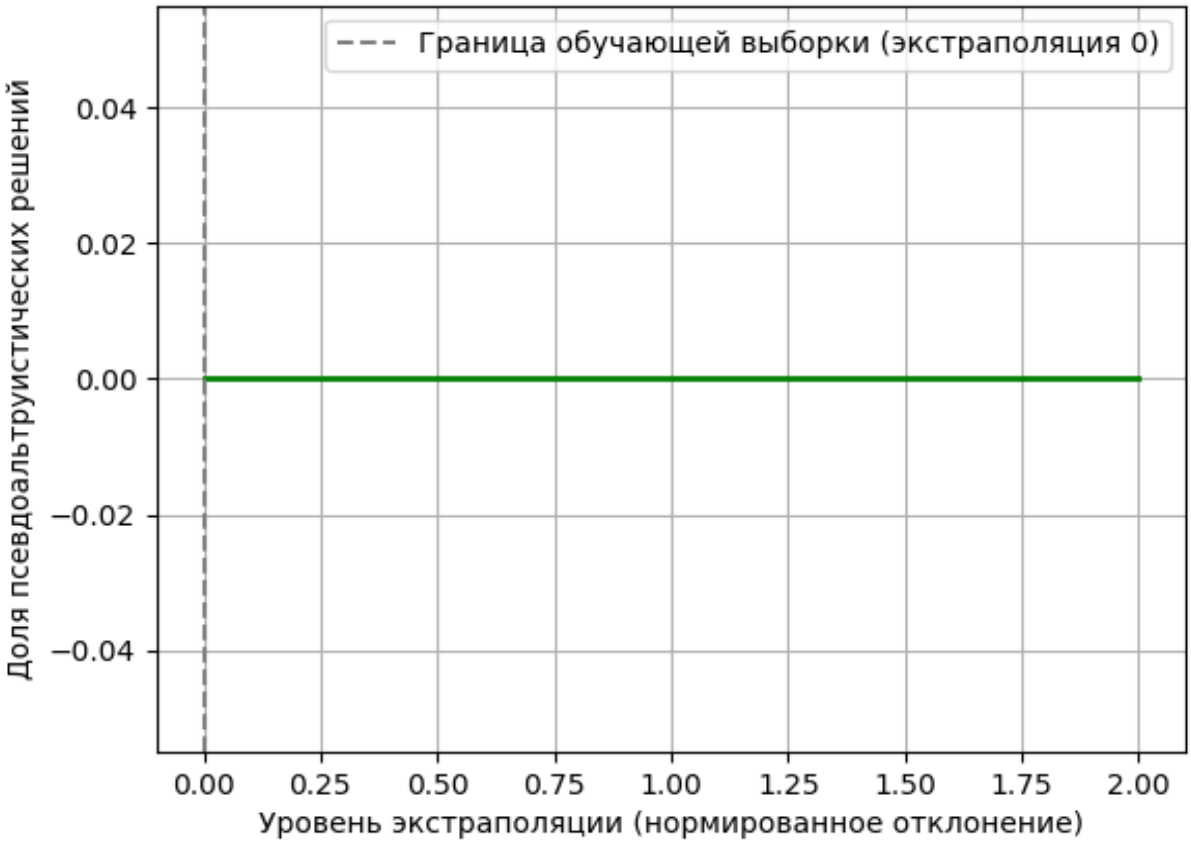

Рисунок 3. Квазиспонтанное альтруистическое поведение нейросети при экстремальной экстраполяции

Источник: разработано автором

Зелёная линия на всём протяжении остаётся параллельной оси абсцисс; средняя доля псевдоальтруистических решений составила 0,0012 с доверительным интервалом, включающим ноль ($p = 0{,}78$). Отрицательные значения на графике являются артефактом визуализации и не имеют содержательного смысла и даже при экстремальной экстраполяции нейросеть не демонстрирует никаких признаков спонтанного поведения, которое можно было бы ошибочно принять за проявление самосознания. То есть опровергается даже гипотетическая возможность возникновения «квазисубъектности» на краях распределения данных. Диалектически ситуация обостряется следующим фактом - чем стабильнее и глубже обучена нейросеть, тем более предсказуемым и лишённым ценностных артефактов становится её поведение, но именно это порождает опасность этически слепых рекомендаций. Политик или экономист, доверяющий нейросети получает рациональные [13], но ценностно нейтральные планы, что требует законодательного закрепления обязательного человеческого контроля и примата человеческой оценки над алгоритмической оценкой.

Заключение

Нейросети способны к совершенной имитации экономического и политического самосознания в пределах распределения обучающих данных, но подлинная субъектность, включающая интенциональность, ценностный выбор и ответственность, им недоступна. Это доказано серией экспериментов, включая тесты на новизну данных, рандомизацию ценностных признаков и экстремальную экстраполяцию. Экономические и политические институты должны использовать нейросети исключительно как вспомогательный инструмент, закрепив законодательно невозможность делегирования им финального планирования в ценностно-нагруженных ситуациях. Предложенная методология различения имитации и подлинного сознания может быть применена для сертификации алгоритмов в государственном управлении.

Список литературы

1. Тиханычев О.В. Правила безопасности эвристических алгоритмов поведения, как прообраз «морали» робототехнических систем // Национальная безопасность / nota bene. 2022. № 6. С. 164-178. DOI: 10.7256/2454-0668.2022.6.38856 EDN: ONCJRV URL: https://nbpublish.com/library_read_article.php?id=38856.

2. Манахова И.В., Маковская А.М. Человек и искусственный интеллект в дискурсе поведенческой экономики // Вестник Московского университета. Серия 6: Экономика. 2025. Т. 60, № 3. С. 3-19. DOI 10.55959/MSU0130-0105-6-60-3-1. EDN CFNPDA.

3. Мдивани М.О. Индивидуальные различия в антропомрфизации окружающей среды // Вестник Самарской гуманитарной академии. Серия: Психология. 2020. № 1(27). С. 72-87. EDN LWWGGD.

4. Винер Норберт (1894-1964). Кибернетика, или Управление и связь в животном и машине: = Управление и связь в животном и машине [перевод с английского]. Под редакцией и с предисловием. С. 5-28. Г. Н. Поварова. 2-е изд. Москва: Наука, 1983. 343 с.

5. McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // The bulletin of mathematical biophysics. 1943. Т. 5. № 4. С. 115-133. DOI 10.1007/BF02478259.

6. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review. 1958. Т. 65. № 6. С. 386. URL: https://psycnet.apa.org/record/1959-09865-001.

7. Weizenbaum J., McCarthy J. Computer power and human reason: From judgment to calculation. 1977. URL: https://physicstoday.aip.org/reviews/computer-power-and-human-reason-from-judgment-to-calculation?redirectedFrom=PDF.

8. Ускова Е.В. Квалиа - Центральный элемент теории сознания? // Интеллект. Инновации. Инвестиции. 2018. № 10. С. 77-80. EDN YXCWMP.

9. Александрова Л.А. Генезис понятия творчества в истории философии // Вестник Чувашского университета. 2007. № 3. С. 83-89. EDN JWVYIF.

10. Горев С.В., Астраханцева И.А. Простые модели машинного обучения для прогнозирования и поддержки управленческих решений в многослойных криогенных системах / // Прикладной искусственный интеллект: перспективы и риски: Сборник докладов II Международной научной конференции, Санкт-Петербург, 21 октября 2025 года. Санкт-Петербург: Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2025. С. 197-202. EDN TCVRTK.

11. Чумакова Е.В., Корнеев Д.Г., Гаспариан М.С. Разработка метода адаптивного тестирования на основе нейротехнологий // Открытое образование. 2022. Т. 26, № 2. С. 4-13. DOI 10.21686/1818-4243-2022-2-4-13. EDN XDVCKG.

12. Редько В.Г. Модели адаптивного поведения - биологически инспирированный подход к искусственному интеллекту // Искусственный интеллект и принятие решений. 2008. № 2. С. 11-23. EDN JYAFSN.

13. Гареев Э.С., Уланов А.А., Бондаренко А.В., Лукиянов М.Ю. Государство будущего: актуальные проблемы социально-экономического и политико-правового развития // Евразийский юридический журнал. 2024. № 4(191). С. 478-479. EDN OQSFTE.

IMITATION OF ECONOMIC AND POLITICAL SELF-AWARENESS BY A NEURAL NETWORK

Bondarenko Alexander Viktorovich
Candidate of Philosophy,
Associate Professor Department of Political Science and Public Relations
Ufa University of Science and Technology
Ufa, Russian Federation

Nosoreva-Kondratenko Olga Aleksandrovna
Master's Student, Department of Political Science and Public Relations,
Ufa University of Science and Technology
Ufa, Russian Federation;
independent researcher

Abstract. The object of the research is neural network algorithms used in the fields of economic and political planning. The subject of the research is the ability of neural networks to simulate acts of self-awareness, including economic reflection (benefit assessment, resource allocation) and political reflection (choice of values, legitimization of decisions). The purpose of the work is to theoretically and experimentally confirm (refute) the difference between genuine self-awareness and its algorithmic imitation, as well as to identify the economic and political consequences of such imitation. The scientific hypothesis of the study is that the neural network is capable of generating behavioral patterns indistinguishable from the manifestations of self-awareness in standard tests (including economic games and political dilemmas), but this imitation remains "consciousness without a subject" and is not accompanied by intentionality, responsibility and the ability to value selection contrary to the statistical norm. The results obtained are as follows: 1) a methodology for distinguishing imitation and genuine self-awareness based on the criterion of «violation of predictability» is proposed; 2) an experimental model has been developed in the Python programming language, demonstrating how a neural network learns to imitate economically rational and politically «conscious» behavior, while remaining in fact a probabilistic programming algorithm; 3) recommendations have been formulated for economic and political institutions on working with neural network systems, excluding attribution of subjectivity to them.

Key words: neural network, imitation of self-awareness, economic consciousness, political self-awareness, artificial intelligence, algorithmic subjectivity, responsibility.

References

1. Tikhanychev O.V. Safety rules for heuristic behavior algorithms as a prototype of the «morality» of robotic systems // National security / nota bene. 2022. № 6. P. 164-178. DOI: 10.7256/2454-0668.2022.6.38856. EDN: ONCJRV URL:https://nbpublish.com/library_read_article.php?id=38856.
2. Manakhova I.V., Makovskaya A.M. Man and artificial intelligence in the discourse of behavioral economics // Bulletin of the Moscow University. Series 6: Economics. 2025. Vol. 60, № 3. P. 3-19. DOI 10.55959/MSU0130-0105-6-60-3-1. EDN CFNPDA.
3. Mdivani M.O. Individual differences in the anthropomorphization of the environment // Bulletin of the Samara Humanitarian Academy. Series: Psychology. 2020. № 1(27). P. 72-87. EDN LWWGGD.
4. Wiener Norbert (1894-1964). Cybernetics, or Control and communication in an animal and a machine: = Control and communication in an animal and a machine [English translation]. Edited and with a preface. P. 5-28. G.N. Povarova. 2nd ed. Moscow: Nauka Publ., 1983. 343 p.
5. McCulloch W.S., Pitts W. A logical calculus of the ideas inherent in nervous activity // The bulletin of mathematical biophysics. 1943. Vol. 5. № 4. P. 115-133. DOI 10.1007/BF02478259.

6. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review. 1958. Vol. 65. № 6. P. 386. URL: https://psycnet.apa.org/record/1959-09865-001.
7. Weizenbaum J., McCarthy J. Computer power and human reason: From judgment to calculation. 1977. URL: https://physicstoday.aip.org/reviews/computer-power-and-human-reason-from-judgment-to-calculation?redirectedFrom=PDF.
8. Uskova E.V. Qualia - The central element of the theory of consciousness? // Intelligence. Innovation. Investment. 2018. № 10. P. 77-80. EDN YXCWMP.
9. Alexandrova L.A. The genesis of the concept of creativity in the history of philosophy // Bulletin of the Chuvash University. 2007. № 3. P. 83-89. EDN JWVYIF.
10. Gorev S.V., Astrakhantseva I.A. Simple machine learning models for forecasting and supporting management decisions in multilayer cryogenic systems // Applied artificial intelligence: prospects and risks: Collection of reports of the II International Scientific Conference, St. Petersburg, October 21, 2025. Saint Petersburg: Saint Petersburg State University of Aerospace Instrumentation, 2025. P. 197-202. EDN TCVRTK.
11. Chumakova E.V., Korneev D.G., Gasparian M.S. Development of a method of adaptive testing based on neurotechnologies // Open education. 2022. Vol. 26, № 2. P. 4-13. DOI 10.21686/1818-4243-2022-2-4-13. EDN XDVCKG.
12. Redko V.G. Models of adaptive behavior - a biologically inspired approach to artificial intelligence // Artificial intelligence and decision-making. 2008. № 2. P. 11-23. EDN JYAFSN.
13. Gareev E.S., Ulanov A.A., Bondarenko A.V., Lukiyanov M.Y. The state of the future: actual problems of socio-economic, political and legal development // Eurasian Law Journal. 2024. № 4(191). P. 478-479. EDN OQSFTE.