



<https://doi.org/10.48417/technolang.2024.02.07>

Research article

GPT Assistants and the Challenge of Personological Functionalism

Andrey Yurievich Alekseev¹   and Ekaterina Alekseeva Alekseeva² 

¹Peoples Friendship University of Russia named after Patrice Lumumba. st. Ordzhonikidze, 3, 115419, Moscow, Russian Federation

aa65@list.ru

²State Academic University of Humanities. Maronovsky Lane, 26, 119049, Moscow, Russian Federation,

eaalekseeva@gaugn.ru

Abstract

The paper discusses whether it is correct to speak of “generative artificial intelligence” – a concept that is not within the scope of AI research. The discussion suggests that it is premature to claim that humans are being replaced by GPT assistants such as ChatGPT in the field of sociocultural digital communication. Personological functionalism, which would justify the replacement of people by machines, is based on the psychofunctionalism of Ned Block, who proves the need to psychologize machine functionalism by introducing “meaning” as a criterion for passing the original Turing test. For personological functionalism, in addition to “meaning” the minimum necessary requirements of the Turing test include “creativity.” The paper shows that GPT Assistants do not pass this creativity test. To demonstrate the inability to pass a Turing test for meaningfulness, the Block machine was modified in a pair of 1978 and 1981 papers by combining the neurocomputer with symbolic versions. For the now further expanded Block test, the argumentation of previous versions is preserved and strengthened, leading to the conclusion that machines like GPT Assistants are not capable of fulfilling either the roles of psychological functionalism or personological functionalism.

Keywords: Generative AI; Complex Turing test; Block test; Lovelace test; Chinese nation; Block's machine; Psychofunctionalism; Personological functionalism

Aknowlegment The paper was prepared within the framework of the state task of the State Academic University of Humanities “Digitalization and the formation of a modern information society: cognitive, economic, political and legal aspects.” Registration number R&D 123022000042-0. Topic code FZNF-2023-0004. Registration number of topic 1022040800826-5-5.2.1;6.3.1;5.9.1

Citation: Alekseev, A. Y., & Alekseeva, E. A. (2024). GPT Assistants and the Challenge of Personological Functionalism. *Technology and Language*, 5(2), 80-99. <https://doi.org/10.48417/technolang.2024.02.07>



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



УДК 1: 004.8

<https://doi.org/10.48417/technolang.2024.02.07>

Научная статья

Персонологический функционализм GPT-ассистентов

Андрей Юрьевич Алексеев¹  (✉) и Екатерина Алексеевна Алексеева² 

¹Российский университет дружбы народов имени Патриса Лумумбы. Российская Федерация, ул. Орджоникидзе, д. 3, 115419, Москва, Российская Федерация

aa65@list.ru

²Государственный академический университет гуманитарных наук. Мароновский переулок, 26, 119049, Москва, Российская Федерация

eaalekseeva@gaugn.ru

Аннотация

В статье рассматривается вопрос о терминологической корректности понятия “генеративный искусственный интеллект”. Дискуссия позволяет сделать вывод о том, что преждевременно утверждать о замене человека GPT-ассистентами (Generative Pre-trained Transformer Assistant – GPTA) в среде социокультурной электронной коммуникации. Персонологический функционализм, обосновывающий замену людей машинами, опирается на психофункционализм Неда Блока, доказывающего необходимость психологизации машинного функционализма путем введения “смысла” в продукты прохождения оригинального теста Тьюринга. Для персонологического функционализма минимально необходимыми составляющими тьюрингового теста являются “творчество” и “смысл”. В работе показано, почему GPTA не проходят креативного теста. Для демонстрации неспособности пройти тьюринговский текст на осмысленность дорабатывается машина Блока путем совмещения нейрокомпьютерной (1978 г.) и символической версий (1981 г.). Для нового расширенного теста Блока сохраняется и усиливается аргументация предыдущих версий: машины типа GPTA не способны исполнять ни роли психологического функционализма – в частности, ни персонологического функционализма – в общем.

Ключевые слова: Генеративный ИИ; Комплексный тест Тьюринга; Тест Блока; Тест Лавлейс; Китайская нация; Машина Блока; Психофункционализм; Персонологический функционализм.

Благодарность: Статья подготовлена в рамках государственного задания ГАУГН «Цифровизация и формирование современного информационного общества: когнитивные, экономические, политические и правовые аспекты». Регистрационный номер НИОКТР 123022000042-0. Код темы FZNF-2023-0004. Регистрационный номер темы 1022040800826-5-5.2.1;6.3.1;5.9.1

Для цитирования: Алексеев А.Ю., Алексеева Е.А. Персонологический функционализм GPT-ассистентов // Technology and Language. 2024. № 5(2). С. 80-99. <https://doi.org/10.48417/technolang.2024.02.07>



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



ВВЕДЕНИЕ

В настоящее время с появлением и повсеместным распространением генеративных предобученных трансформеров (Generative pre-trained transformers), компьютерные генеративные модели стали одним из самых заметных явлений современной электронной культуры. Им приписывается едва ли не революционный характер, их внедрение в различные сферы человеческой деятельности выглядит неизбежным, они спровоцировали новый виток научных и общественных дискуссий по поводу искусственного интеллекта. GPT-модели стали обозначаться словосочетанием “генеративный искусственный интеллект”, и это понятие получило признание в том числе и на правительственном уровне. И действительно, GPT-программы продуцируют тексты, изображения, музыку и многое другое, зачастую практически неотличимое от результатов человеческой деятельности, представленных в компьютерном формате. В силу этого возникает соблазн приписывать личностные характеристики и способности GPT-ассистентам – сервисным программам-помощникам, построенным на основе генеративных технологий различного рода. Подобное “персонифицирование” программных продуктов инициирует чрезвычайный интерес к ним. Из-за ажиотажа вокруг этих систем из поля зрения исчезают их очевидные недочеты: недостаточная глубина контекста, зависимость от качества обучающих данных, продуцирование “фейков”, “галлюцинаций”, прочей заведомо ложной информации и многое другое.

Разработка и массовое внедрение сразу несколькими технологическими компаниями таких систем стали возможны благодаря тому, что существенно возросла вычислительная мощность дата-центров, появились огромные массивы открытых данных, которые были сгенерированы пользователями, а затем по сути присвоены генеративными моделями в процессе обучения, что и дало такой заметный и неожиданный эффект. Однако при ближайшем рассмотрении GPT-системы не представляют собой никаких особых “прорывных” программно-аппаратных новаций. Они мало соответствуют приписываемым им свойствам “осмысленного” и “творческого” продуцирования текстовых или мультимедийных артефактов. По своим вычислительным принципам эти модели восходят к марковским цепям и байесовским сетям. В целом же такие модели являются “стохастическим попугаем”: “системой случайного соединения последовательностей лингвистических форм, которые присутствовали в обширных обучающих данных, в соответствии с вероятностной информацией о том, как они комбинируются, но без какой-либо ссылки на их значение” (Bender et al., 2021).

В данной работе мы ставим вопрос о правомерности использования понятия “генеративный искусственный интеллект”, о том, соответствуют ли в принципе GPT-ассистенты научным критериям исследований искусственного интеллекта. Проблема путей дальнейшего развития общества в контексте бума генеративных технологий не изучается. Об этом много других, как правило, восторженных работ.



ГРТ-АССИСТЕНТЫ И ПЕРСОНОЛОГИЧЕСКИЙ ФУНКЦИОНАЛИЗМ

Современные ГРТ-ассистенты (ГРТА) часто демонстрируют персонологические параметры, которые принято считать характеристиками человеческой личности: разум, понимание, мотивацию, намерение, убеждение, смысл, творчество, мораль, красоту, целостность, субъектность и мн. др. Например, в текстах, продуцируемых с помощью данной технологии, помимо присутствия в них искомого пользователем определенного содержания, обнаруживаются лингвистические маркеры эмоциональных состояний, оценочных позиций, включая самооценку, параметры мотивации и т.п. Такие системы, в особенности при дополнении их мультимедийными аудиовизуализациями “личности”, способны сыграть роль актера коммуникации, вводящего в заблуждение взаимодействующих с ним людей. Обман при «подмене личности» средствами, подобными ГРТ, становится общим местом для обозначения рисков и угроз со стороны “генеративного ИИ”.

Выявление персонологических маркеров служит поводом для заявлений в СМИ, на научных и бизнес-конференциях о том, что ГРТ-системы встали на путь “сильного ИИ”. Декларируется, что дальнейшее развитие этих технологий может привести к формированию у них человеческих характеристик: креативности, способности к целеполаганию, принятию решений, моральному вменению и т.п. Здесь же возникает и вопрос о замене людей системами ИИ. Если для замены традиционными системами ИИ рассматриваются профессии с учетом работы в слабо формализуемых областях, но с опорой на автоматизацию рутинных процедур, то в случае с генеративным ИИ заменяются творческие, принципиально неформализуемые профессии. Вопросы приписывания персонологических феноменов результатам функционирования систем ИИ, оценка доверия к процедурам такого приписывания, начиная от дефиниций и выбора их компьютерной архитектуры, изучается в методологии, получившей в последние десятилетия название “функционализм искусственного интеллекта”. Эта методология лаконично представлена в (Алексеев, 2019) на уровне концептуальной организации современных проектов общего искусственного интеллекта: искусственной жизни, искусственного сознания (мозга), искусственной личности, искусственного общества, искусственного мира.

Функционализм ИИ – это направление научно-теоретических и инженерно-технологических исследований, ориентированное на построение и применение компьютерных репликаций (имитаций), репрезентаций (моделей) и репродукций когнитивных феноменов самого широкого спектра жизненных, психических, личностных и общественных проявлений. Исходной основой функционализма ИИ является психофизиологическая теория машинного функционализма Хилари Патнэма 60-х гг. прошлого века. В 1960 году он утверждал, что при изучении интересующих вопросов ЭВМ может выполнять роль, инвариантную функциям мозга (Putnam, 1960).



Например, оценивать боль другого человека можно трояко: во-первых, бихейвиорально, по стонам; во-вторых, физикально, по изучению паттернов нейтральной динамики; в-третьих, функционально, по отказу “триггера № 36” в чипе, который репродуцирует нейроны, ощущающие боль, в составе работы мозга как системы. В 1997 году автор машинного функционализма отказался от собственных идей, указав на сложность выявления функциональных связей и в обществе, и в человеке: “Когнитивная физика невозможна, точно так, как ранее стала очевидной невозможность социальной физики О. Конта” (Putnam, 1997). Однако вряд ли это утверждение способно остановить сторонников современного функционализма ИИ. Сегодня в формировании функционализма ИИ выделяются собирательный, определительный и наблюдательный подходы (Алексеев, 2019). Собирательный подход отвечает за сбор, идентификацию, координацию, формализацию, систематизацию, унификацию, кодификацию всевозможных функционалистских теорий. Определительный подход обеспечивает анализ и выявление главных функционалистских характеристик, отношений, закономерностей, причинностей, инвариантных относительно содержания когнитивных феноменов. Наблюдательный подход позволяет оценивать различные отношения человека к компьютерному миру систем ИИ: онтологические, эпистемологические, логические, лингвистические, аксиологические, эстетические, этические, праксеологические и другие фундаментальные отношения.

Возникает вопрос: как вписывается идея “генеративного ИИ” в методологию современного функционализма ИИ, в обозначенные выше и фундаментальные отношения? И вообще вписывается ли она в эти подходы?

Машинный функционализм семидесятилетней давности апеллировал к идее машины Тьюринга. В те годы тест Тьюринга не достиг популярности и про статью (Turing, 1950) упоминали единицы. Поэтому Х. Патнэм в (Putnam, 1960) идею теста Тьюринга просто-напросто не заметил. Современный функционализм ИИ базируется на методологии комплексного теста Тьюринга, который систематически изучает принципиальные вопросы компьютерной реализации когнитивных феноменов самого широкого спектра. В (Алексеев, 2008, 2013, 2016; Лекторский, 2022) рассматривался “костяк” наиболее популярных тьюринговых тестов, отвечающих на вопросы: может ли машина, как человек, мыслить, сознавать, понимать, творить, любить, дружить, быть личностью, свободной, добродетельной и пр.

На первый взгляд кажется, что персонологический функционализм, изучающей вопросы “замены” естественных личностей функциональными “двойниками” задается тестами, которые близки проблематике компьютерного моделирования личности. Например, тест “искусственной личности” (достаточно подробно раскрыт в (Алексеев, 2016)) и тест “зомби” (Алексеев, 2020) воспроизводит метафизическую драму персонажей мысленных экспериментов философии сознания. Д. Деннетт полагает, что все мы, люди, на сегодняшнем этапе антропосоциогенеза, являемся всего лишь зомби, но можем стать сознательными в условиях глобального ИИ, который позволит четко осознать роль и место каждого



в системе человечества в целом. Дж. Серль, напротив, убежден, что мы в далеком прошлом обладали полнотой личностных способностей, а вот ИИ-системы ускоряют деградацию людей и скоро мы превратимся в зомби.

Эти метафизические картинки забавны, и при этом они полезны, так как выражают логически достаточные основания того, чтобы GPT-система проходила тест на “личностное”. Имеется много иных когнитивных феноменов, которые необходимо имитировать средствами GPT – понимание, подсознание, эмоция, самоопределение, воля... На наш взгляд, логически необходимыми для выявления функциональных персонологических инвариантов является “смысл” и “творчество”. Если нет этих феноменов, то ни о каких проявлениях “личности” утверждать нельзя, это просто-напросто логически невозможно.

ПРОХОДЯТ ЛИ GPT-АССИСТЕНТЫ ОРИГИНАЛЬНЫЙ ТЕСТ ТЬЮРИНГА?

GPTA проходит оригинальный тест Тьюринга, игру в имитацию интеллекта. Вопрос о прохождении тьюринговой игры в имитацию, как правило, предшествует другим вопросам любого частного теста из состава комплексного теста Тьюринга. Следует сразу отметить, что эта “игра” выходит за границы методологии бихейвиоризма: тьюринговые “диспозиции” вычислимы. Вычислимость трактуется не как арифметические преобразования, а как порядок следования программе. Этот порядок в общем случае кодируется целыми числами. В бихейвиоризме предметом когнитивного анализа выступает не некая ментальная структура сознания, а объективно фиксируемые параметры поведения (реакции), определяемые внешними воздействиями (стимулами). В лингвистическом бихейвиоризме стимулами и реакциями диспозиций являются вербальные выражения, и так как осуществляется манипуляция семиотическими комплексами, то не получается игнорировать ни семиотическую, ни прагматическую компоненту тьюринговой коммуникации.

Первая компонента предполагает отношение символьных выражений к объективной, субъективной или intersубъективной формам реальности, т.е. выделение ментального параметра как *значения*. Вторая компонента определяет отношение носителя выражений к выражениям с учетом бесчисленного контекста познавательных, производственных, экономических, социальных, культурных, моральных и прочих отношений. Получается, что в диспозициях должен присутствовать *смысл* как специальный предмет программирования. Оценка успешности компьютерной имитации интеллекта осуществляется судьей и зависит от его наблюдательности, компетентности, ума, бескорыстности и пр. Судья вправе присудить игрокам не только правильный гендер, как изначально задумывал автор оригинального теста, но и наличие чувств, мотивов, воли и многого другого, в том числе и личностного. Конечно, необходим некоторый минимум компьютерных достижений имитации интеллекта, по крайней мере, способности вести диалог. Способен ли лингвистический бихейвиоризм, научившийся программировать диспозиции и, благодаря этому, преступивший порог функционализма,



имитировать креативность и осмысленность ответов на вопрос? Для оценки этого в составе комплексного теста Тьюринга изучаются тест Лавлейс и различные версии теста Блока.

СПОСОБНЫ ЛИ GPT-АССИСТЕНТЫ ПРОЙТИ ТЕСТ ЛАВЛЕЙС?

GPTA не проходит тест Лавлейс, т.е. тест на креативность. Впервые “тест” на компьютерное творчество предложила Ада Лавлейс в 1843 г.: машина способна на многое, например, вычислять, исполнять музыкальные симфонии. Но при этом способна делать только то, что ей указывает (инструктирует) сделать человек. Машина творить не может! (A.L.L., 1843).

А. Тьюринг в 1950 году сделал послабление в дефинициях: творческим является то, что удивляет наблюдателя и является неожиданным. Например, благодаря машинному обучению человек лишь предполагает общий ход продуцирования артефакта, конкретное же воплощение достигается в функциях (нейро)компьютера (Turing, 1960, p. 469). Но такая конкретика, по мнению А. Тьюринга, весьма необычна и её сложно предсказать.

В 2000 г. ряд авторов предложили улучшенный тест Лавлейс, в котором фигурировали “артефакты”, благодаря которым – подчеркнем это – GPT-системы стали популярны – картины, поэмы, музыка (Bringsjord, Bello, Ferrucci, 2001). Однако способна ли GPTA пройти тест Лавлейс? Компьютер пройдет этот тест тогда и только тогда, когда выполняется система аксиом: 1) GPTA способен продуцировать артефакт *A*; 2) артефакт *A* не является результатом обмана, случая или сбоя; 3) GPTA всегда может *A* репродуцировать (повторить); 4) судья *I* не способен объяснить, как GPTA продуцировал *A*, несмотря на полное понимание механизма GPTA. Предложенная аксиоматика приводит к следующим соображениям: 1) Креативный тест невозможно пройти путем обмана. 2) Тест Лавлейс всем современным GPTA присваивает статус некреативных. Модели запрограммированы разработчиками на основе больших данных, ничего “от себя” такие GPTA добавить не могут. 3) Необходимым условием прохождения теста Лавлейс является способность GPTA к программированию собственного программного кода, новых аксиом (для символического подхода), нового способа обучения нейронной сети (для коннекционистского подхода). Но это невозможно из-за аргумента Геделя-Лукаса: механическая система, которой является GPTA, не способна к аутоформализации. Для GPT даже отдаленно не стоит вопрос о том, способна ли программа порождать другие программы методами, скажем, “автогенеративного программирования”. Принцип работы GPTA принципиально исключает аксиоматику как вычислительную основу функционирования. Сегодня часто встречается утверждение о том, что GPT мгновенно формируют многообразие разных программ по требованию “промптов”. Однако на самом деле новые программы не создаются: GPT предлагает фрагменты кода, которые были ранее представлены в Интернете. Моделирование “свободной воли”, имитация мотивационно-волевых механизмов человека, стремящегося к созданию чего-то нового, дизрезультативно, так как продукт представляется в некотором



пространстве потенциально представимых решений. GPT-генерация не доходит до уровня аксиоматического представления этого пространства. Тест Лавлейс собственно и показывает невозможность компьютерной реализации феномена творчества. Однако возможна интерпретация продукта компьютерной генерации как квазикреативного, “как бы” творческого продукта, которая определяется компетентностью тьюрингового судьи, приписывающего программе креативность.

Чтобы повысить экспертные полномочия тьюрингового судьи в 2014 году М. Ридл в “тесте Лавлейс 2.0” (Riedl, 2014) предложил новую версию исходных предпосылок, которую применительно к нашему случаю представим таким образом: 1) GPTA продуцирует артефакт A типа T ; 2) A должен соответствовать набору ограничений C , где $c_i \in C$ – любой критерий, выражаемый на естественном языке; 3) человек-оценщик H , выбравший T и C , удовлетворен тем, что A является допустимым экземпляром T и соответствует C ; 4) человек-судья I определяет комбинацию T и C как нереальную для обычного человека. То есть помимо человека, выполняющего роль тьюрингового судьи, появляется человек (H), компетентно оценивающий гениальность, талантливость, креативность артефакта A или, напротив, бесталанность, обыденность этого произведения. Произошла четкая спецификация критериев оцениваемого продукта. Человек-оценщик и человек-судья должны быть экспертами в достаточно ограниченной области, т.е. нужен конкретный специалист, а не “человек вообще”, определяющий, может или не может творить машина. Самое главное, продукты генерации GPTA должны быть осмысленны.

В 2015 году Т. Пожарев предложил *тест Лавлейс 3.0*, когда для постижения смысла оценщик H должен выявлять факт осмысленного продуцирования артефакта A (Алексеев & Пожарев, 2020). Например, какая операция использовалась при создании A : абстрагирование или обобщение? Если оценщик H обнаружил абстрагирование, то возникает повод утверждать креативность. Если H обнаружил обобщение, то это повод отрицать креативность. Нейронные сети не формируют идеальные объекты, не оперируют ими, они способны только обобщать и ограничивать. Формально-логически это означает, что GPTA работает исключительно с объемами понятий, экстенционально. С содержанием понятий, со смыслами, интенционально, работает исключительно человек.

ПРОХОДЯТ ЛИ GPT-АССИСТЕНТЫ ТЕСТ БЛОКА?

GPTA не проходит тест Блока, т.е. тест на осмысленность. Тест Блока входит в состав комплексного теста Тьюринга и является совершенным частным тестом на “глобальный” искусственный интеллект. Тест Блока был сформулирован в двух работах Неда Блока (Block 1978, 1981) почти полувековой давности, но сегодня оказался чрезвычайно востребован в связи с GPT-экспансией. В ней представлены две версии проекта вычислительной машины для работы со “знаниями” всего человечества: коннекционистская и символическая. Одна выражает идею глобализации знаний в среде Интернета, вторая версия обеспечивает ответы по запросам, которые сегодня называются “промтами”. Первая версия,



коннекционистская (точнее, нейросетевая), предложена в статье Н. Блока 1978 г. “Проблемы с функционализмом”. Этот проект автор назвал “Китайской нацией”. Сегодня он иногда именуется “Китайский мозг” (“China brain”) (Bringsjord et al., 2001). В философии ИИ более известным является другой “китайский проект”, появившийся два года спустя – “Китайская комната” Джона Серля 1980 г. Комментаторы отмечают (например, (Cole, 2023)), что на серлевский мысленный эксперимент повлияла отмеченная нами работа Н. Блока (Block, 1978). Коннекционистская машина Блока масштабирует когнитивные функции психики отдельного индивида до уровня т.н. “общественного сознания”. На выбор названия повлиял размер населения Китая, сравнимый с количеством нейронов в коре головного мозга человека. Предлагается идея связать паттерны активности мозга китайца, которого избивают, с паттернами активности китайской нации в целом, когда каждый активный нейрон китайца-индивида активизирует нейрон китайца-члена общества. Если срабатывает нейрон, то через спутник Земли посылается сигнал на мобильное устройство члену китайской нации, скажем, прыгнуть. Если нейрон дезактивируется, то поступает сигнал – замереть. В принципе, возможен функциональный инвариант между нейтральной динамикой мозга избиваемого китайца и китайской нацией. Ведь количество китайцев и количество нейронов сравнимо. Но будет ли нация испытывать боль, аналогичную боли китайца-индивида? На первый взгляд очевидно отсутствие каких-либо корреляций. Однако так ли это? Вопрос до конца не разрешим, как и вопрос об эпистемологическом реализме общественного сознания, ноосферы и глобалистики в целом. Конечно, тест Блока “Китайская нация” поднимает мировоззренческие проблемы, которые возникают в мысленных экспериментах в весьма далекой от реалистической презумпции лингвистической философии.

Вторая версия машины Блока, символическая, представлена в статье 1981 г. “Психологизм и бихейвиоризм” (Block, 1981). Она практичнее первой версии и позволяет изучать концептуальные модели построения глобальных интеллектуальных технологий. Эта версия называется по-разному: по-кибернетически звучит как “черный ящик” (“Black Block”). Иногда называют “голова-ящик”, иногда, ссылаясь на фамилию автора, “головой Блока” (Blockhead (Ned Block, 2024)). Н. Блок предлагает вообразить традиционную информационную, не нейрокомпьютерную, машину представления, хранения, обработки и выдачи ответов на все возможные вопросы. Глобальная машина собирает все эти лингвистические диспозиции, систематизирует их и по сути реализует поиск подстрок, так как в начало запроса включается дескриптор “диспозиции” в виде вопроса. Остается найти вопрос в глобальной базе данных и часть оставшейся подстроки подставить как ответ пользователю, который задал вопрос. Принципы работы машины поиска подстрок мало чем отличаются от общих способов функционирования GPT. Получается то, что сегодня реализуется поисковыми сервисами Интернета. Правомочен вопрос: является ли машина поиска подстрок системой искусственного интеллекта?

По отдельности эти версии машины Блока не подходят для концептуального анализа GPT-систем. Первая ориентирована на решение метафизических вопросов



философии сознания: можно ли представить в компьютерной системе феномены общественного сознания и каковы перспективы компьютерной реализации этого представления? Вторая версия не лишена практической ориентации, например, программирования интеррогативных (вопрос-ответных) сценариев использования семантических сетей на основе всех возможных открытых баз данных и знаний. Однако если две эти машины совместить в едином проекте, то получится вполне правдоподобная символно-коннекционистская машина реализации теста Блока. Назовем ее GPT-машиной Блока или *новой машиной Блока*. Эта машина учитывает традиционные для машинного обучения особенности построения GPT-систем: дискретизация “текстовых диспозиций” в “токенах” и хранение их цифровых аналогов в “мешках слов”, определение ассоциаций кодов и формирование пространства векторов, учет статистики частот встречаемости ассоциаций.

Попытаемся осуществить следующие аналогии в сфере методологии науки. Машинный функционализм концептуально реализуется индетерминированной машиной Тьюринга, психофункционализм – машиной Блока, а персонологический функционализм – новой машиной Блока. Психофункционализм сегодня сложился как самостоятельное направление в аналитической философии (Levin, 2023), в философии искусственного интеллекта в целом и в функционализме ИИ в частности. Психофункционализм дополняет идею функционализма Х. Патнэма: программа имитации психических феноменов, бихейвиорально реализующая реакции системы (выход) на ее стимулы (вход), необходимо включает психологические параметры. Как мы ранее отмечали со ссылкой на Х. Патнэма, боль человека представляется “*включением триггера № 36*” как коррелята соответствующих мозговых стимулов (Putham, 1960, p. 363). Помимо этого, согласно Н. Блоку и исходя из психофункционалистской интерпретации психофизического функционализма, в состав “теории боли” должна входить “*теорема № 17*”, которая на внутреннем языке мысли представляет “предикатор боли», определенный на “входе” из значений переменных “стол», “гвоздь», “сесть» и на “выходе” со значениями переменных “кровь”, “крик”, “стон” (Block, 1978, p. 236). Физикализм в этом построении опосредованно участвует как функциональное тождество физически различных структур относительно компьютерного блока (*триггера № 36*), мозга человека, китайской нации, тины марсианина и пр., коррелированных с психологической *теоремой № 17* и с субъективно воспринимаемым феноменом “*боли*” от первого лица или с эмпатическим переживанием боли относительно оцениваемых субъектов (интерсубъективно, от второго лица).

Значимость психофункционализма обосновывается в следующем порядке. Вначале выдвигается как бы вполне очевидный тезис о том, что тест Тьюринга – разновидность бихейвиоризма. Тезис разрушается под натиском стандартных антибихейвиористских аргументов. Далее за счет нестандартных антибихейвиористских аргументов уничтожаются остатки тех нюансов лингвистического бихейвиоризма, которые привносит тьюринговая игра в имитацию. Программистские патнэмовские расширения лингвистического



функционализма разрушаются, то есть выявляется несостоятельность машинного функционализма и подчеркивается истинность психофункционализма.

К стандартным антибихейвистским аргументам, применимым для обоснования психофункционализма GPT-ассистентов, относится, во-первых очередь, *аргумент Чизхолма-Гича* (Block, 1981, p. 11) совмещающий идеи против бихейвизма при восприятии (выдвинуты Родериком Чизхолмом (Chisholm, 1957) и, в общем, ментальных актов (изучены Петером Гичем (Geach, 1957)). Эти идеи утверждают невозможность абстрагированного выделения в поведенческой диспозиции некоторого определенного психического состояния без учёта всей совокупности иных психических состояний. Допустим, бихейвист исследует желание человека съесть мороженое. “Желание” представляется в форме поведенческой диспозиции. В структуру этой диспозиции входит, например, намерение немедленно схватить мороженое, особенно если его дарят (“пока не передумали”). Однако возникает конфликт интересов. Возможно, что человек сомневается в том, что ему предлагают именно мороженое, а не муляж мороженого, предлагаемый шуток ради. Нужна так же уверенность в том, что за этим не последует противоречия с другими более важными желаниями. Не нужна ли ответная услуга, которая может оказаться непосильной? Таким образом, входы и выходы диспозиции изменяются весьма непредсказуемым образом. Относительно “боли” возможны следующие диспозиции, предложенные Х. Патнэмом. Возможны различные комбинации психических состояний, таких как: [боль + обычное переживание боли] или [отсутствие боли + желание обмануть, что боль переживается]. Во втором случае правдоподобны *суперспартанцы Х. Патнэма*, которым законы запрещают выказывать боль несмотря на ее испытывание. Правдоподобен *совершенный актер*. Он может успешно притворяться в том, что ему больно, хотя на самом деле это не так. Паралитики и Мозги в бочке не могут выказать боль, так как им нечем это осуществить.

На основе этих аргументов, но уже применительно не к боли, а к мышлению, Н. Блок утверждает: интеллектуальное поведение продуцируется комбинацией: [интеллект + склонность к мышлению]. Невозможны комбинации: [отсутствие интеллекта + желание обмануть в наличии интеллекта] или [отсутствие интеллекта + желание казаться разумным]. Умный может симитировать глупость, но глупый не способен симитировать разум – полагает Н. Блок.

Этих аргументов достаточно, чтобы разрушить лингво-бихейвистскую концепцию игры в имитацию. После проведенной критической деконструкции Н. Блок формулирует новый тест Тьюринга: диалоговый интеллект, проявляющийся в разговоре людей – это способность продуцировать осмысленную последовательность вербальных реакций на некоторую последовательность вербальных стимулов, какими бы эти стимулы не были (Block, 1981, p. 18). Не совсем понятен логико-экзистенциальный статус последней части утверждения. Мы интерпретируем их в формате “смысловой индефферентности”: не важно, являются ли осмысленными либо бессмысленными входные стимулы. Главное, чтобы результат был осмысленным или был похожим на осмысленный.



Тест Блока как новый тест Тьюринга, *осмысленный*, а не случайно сгенерированный, снимает возражение Чизхолма-Гича: возможно сколь угодно много вариантов диспозиций, убеждений, намерений, желаний тому, что GPTA не будет расположен выдавать осмысленные ответы. Тем не менее, разнообразные варианты диспозиций никак не сказываются *на способности* GPTA формировать такие ответы. Но такие способности не являются операциями машины. Это суть способности *людей* понимать стимулы и формировать реакции как осмысленные тексты. Снимаются аргументы “суперспартанец” и “совершенный актер”: надо обладать недюжинными интеллектуальными способностями для того, чтобы в совершенстве имитировать отсутствие разума. Невозможно притворяться разумным и не быть таковым. Для интеллектуалов-паралитиков и “умных мозгов в бочке” вообще нельзя найти опровергающие примеры – у них, возможно, имеется способность реагировать осмысленно, однако для реализации этой способности им не хватает средств, которые могут выразить способность постигать “смысл”. Теоретически “смысл” можно объективировать, например, вершиной треугольника Фреге. На практике же смысл, не являясь значением, значит. Это смысловое схватывание требует субъекта, владеющего и оперирующего смыслом, раскрытия внутреннего мира человека. Невозможно вообразить, что бессмысленная комбинация символов, демонстрирующая отсутствие интеллекта при ее производстве, способна, минуя человека, механистично, но осмысленно отвечать на произвольную последовательность стимулов.

Однако откуда в GPTA возможно продуцирование осмысленных реакций? Для “отражения” нестандартных антибихейвиористских аргументов Н. Блок кратко описывает, но детально раскрывает принцип работы машины, которая на уровне всего человечества способна производить осмысленную последовательность вербальных реакций на вербальные стимулы. Это так называемая машина поиска подстрок. Задача машины – по ответу на запрос сформировать поиск подстроки, представленной как фрагмент строки, которая была ранее сформулирована, отформатирована и сохранена в базе знаний диалоговых сценариев всего человечества. Машина Блока представлена почти полвека назад, в 1981 г., при этом она предвосхищает принцип работы GPTA.

Мы не будем выделять очевидные отличия, они прозрачны. Подчеркнем то, что машина Блока на поведенческом, бихейвиорально оцениваемом уровне выглядит как интеллектуальная. Однако знания её внутреннего устройства убеждают в полном отсутствии у неё “интеллекта”. Вся совокупность осмысленных вербальных выражений (которая к тому же является конечным множеством) на выходе такой машины задаётся исключительно человеческим коллективом (проектировщиков, инженеров, программистов и др.). Люди долго и упорно трудятся для реализации всевозможных последовательностей осмысленных ответов на предполагаемые последовательности вербальных стимулов. В ходе работы используются специальные инструментальные средства, например, средства автоматизации программирования, машинного обучения для предварительной разметки траекторий ответов, средства ранжирования статистики частоты встречаемости подстрок. Самое главное – люди творчески воображают и



решают, что считать осмысленной последовательностью языковых выражений, а что является бессмыслицей.

Такая машина, несомненно, является разумной. Однако интеллект, который машина “проявляет” как в форме актуального поведения, так и в форме диспозиции интеллектуального поведения – *суть интеллект людей, которые создают тексты, формулируют вопросы, осуществляют поиск ответов и при обнаружении неудовлетворительных ответов отрицают или исправляют*. То есть это система естественного интеллекта, но отнюдь не искусственного интеллекта. Машина Блока – это как двухстороннее радио, радиоприемник в одну и в другую стороны акторов диалога (Block, 1981, p. 21-22), которому абсурдно приписывать способность мышления и понимания. Столь же абсурдно приписывать “интеллект” компьютеру, если он исполняет роли, инвариантные функциям радиоприемника. Если теперь воспользоваться первой, нейрокомпьютерной версией машины Блока и соединить со второй, то получим функциональную эквивалентность GPTA с радиоприемником. Поэтому вся аргументация неинтеллектуальности машины Блока, тщательно разработанная для первой версии машины, переносится и для случая с GPTA.

Мы полностью согласимся с этой объемной аргументацией, указав на ряд интересных моментов, сохранив, по возможности, порядок последовательности аргументов, как было в (Block, 1981).

Первый абзац аргумента отражает аргумент Блока. Второй абзац – опровержение этого аргумента. Если нет концептуальных коллизий между GPTA и машиной Блока, в этом списке применяется аббревиатура анализируемой нами машины, т.е. “GPTA”.

1. *Чрезмерная антропологизация системы ИИ*. Интеллект, проявляемый машиной Блока, является интеллектом разработчиков – программистов, экспертов, операторов, инженеров по знаниям и др.

Машина не обнаруживает отличий между поведением, которое отражает собственный интеллект машины, и поведением, отражающим только интеллект разработчиков машины.

2. *Актуальность знаний*. В машине Блока содержатся только те знания, которые в нее введены. Принципы работы GPTA принципиально отличаются от классических принципов ИИ. В них нет гипотез, абдукции, индукции, традукции. В машине только факты, которые могут оказаться фейками.

Систему следует считать интеллектуальной, даже если в неё не заложены знания о текущих событиях.

3. *GPTA не является интеллектуальной системой*. GPTA как машина с внутренней механической структурой не интеллектуальна даже тогда, когда выглядит интеллектуальной в любом внешнем отношении. Система считается интеллектуальной в той степени, в какой её действия имитируют поведение действительно интеллектуальной системы. При этом игра внешне интеллектуальной системы ошибочно принимается за признак её интеллектуальности.



Так как операции машины Блока суть полностью имитация вопросно-ответных сценариев, то они не могут служить основанием для приписывания машине интеллекта.

4. *Интеллектуальный шовинизм.* Возможно, некоторый “шминтеллект” интеллектуальнее “интеллекта” в силу различия и даже несопоставимости информационных процессов для их реализации.

Так как внутреннее строение GPTA убеждает в том, что она успешно имитирует игру в имитацию интеллекта, не обладая интеллектом, то этого контрпримера, по мнению Н. Блока, достаточно для понимания того, “интеллект» – это нечто иное, нежели чем то, что имитирует интеллект.

5. *Легитимность машинного интеллекта.* Вообразим, что некто, обладающий позитивным правом, постановил считать, что человеческий интеллект – это машинная обработка информации точно такая же, как это осуществляет GPTA. Возможно так же, по утверждению позитивно-правового деятеля и то, что человек менее интеллектуален и, следовательно, менее интересен, нежели чем машина. Он же не имеет общечеловеческого масштаба хранимых данных, параллелизма, скорости вычислений. Можно ли утверждать, что люди менее разумны, нежели чем GPTA?

Подобного рода “законотворческие постановления” расходятся и со здравым смыслом, и с исследованиями когнитивной науки. GPTA не интеллектуальна, поэтому позитивное право приписывания интеллекта неинтеллектуальной машине, если и установится, то не надолго.

6. *Тождество мышления и GPTA.* Человек мыслит так же, как мыслит машина Блока (GPTA). Собирает знания на протяжении всей жизни и вероятностно отвечает на вопросы.

Когнитивные процессы, без всякого сомнения, гораздо более механистичны, чем это принято считать. Однако огромна разница между утверждением, что люди более механистичны, и утверждением, что человек – это машина Блока.

7. *Комбинаторный взрыв машины Блока.* Является ли эмпирической гипотезой то, что интеллект – это способность выдавать осмысленные последовательности реакций, соответствующие входным последовательностям? Машина Блока невозможна из-за комбинаторного взрыва.

Существует порядка 10^{30} грамматических предложений длиной в 20 слов (Н. Блок ссылается на классическую книгу по когнитивистике Дж. Миллера, Е. Галантер и К. Прибрама (Miller, Galanter, & Pribram, 1960)). Предположим, что из них 10^{15} семантически корректны. Тесту Тьюринга длительностью в один час может понадобиться порядка 100 таких предложений. А это 10^{1500} строк – число, которое больше числа частиц во вселенной.

8. Однако проблемы комбинаторного взрыва преодолены GPTA. Пожалуй, одним из достоинств современного уровня вычислительной техники является суперкомпьютерная предобученная схема выдачи ответов на запросы.

В самом деле, для GPTA нет проблемы с этим взрывом. Машина обучается на последовательностях текстов, которые ранее уже были осмыслены и представлены человеком. Самостоятельно GPTA ничего осмысленного не формирует.



9. *Редуцируемость интеллекта к диалоговому интеллекту.* Насколько правомерна эмпирическая гипотеза о том, что интеллект – это способность выдавать осмысленные последовательности реакций, соответствующие входным последовательностям?

Отождествление интеллектуальности со способностью выдавать осмысленные ответы – это второстепенный принцип эмпирической психологии. Подобного рода отождествление может быть предложено как рациональная реконструкция неопределенно общей концепции интеллекта, которая, возможно, найдёт полезное применение в будущей теории эмпирической психологии. Люди со здравым рассудком могут неправильно понимать предложения, путаться. Любому же нормальному человеку, который участвует в длительном тесте Тьюринга, вскоре должно наскучить это мероприятие, он станет невнимателен. Говоруну с самого начала будут нести бессмыслицу, иногда извиняясь, что не слушали собеседника. Многие люди на замечания собеседника будут отвечать свободными ассоциациями, нежели напрягать свой разум, давая вразумительные ответы. Некоторые будут постоянно жаловаться на непривлекательность этих бесконечных тестов Тьюринга.

10. Самым лучшим проектом интеллектуальной машины станет *замена памяти разумом*. “Для теста Тьюринга неопределённой длины машина становится настолько большой, что попытка ее построить станет причиной провала средств финансирования проекта в черную дыру” (Block, 1981, p. 34).

Техническая изобретательность, сколь бы изощрённой она не мыслилась, выходит за рамки методологии моделирования речевых или диалоговых способностей человека. Эвристики недостаточно для этого.

11. *GPTA и аниматроника.* Лишь немногие люди смогут пройти тест Тьюринга, который длится девяносто лет, и ни один человек не сможет пройти тест Тьюринга длиной в пятьсот лет (Block, 1981, p. 34).

Длина машины Тьюринга не является необходимым условием интеллектуальности. Лучше сравнить машину Блока с печально известными “машинами” прошлых столетий, в которых незаметно прятался человек, слушал вопросы и отвечал на них посредством механизмов. Казалось, что отвечает машина, но на самом деле отвечал человек. GPTA вполне способна играть роль такой “машины”. Она хранит “знания” людей, ничего не добавляя от себя.

12. *GPTA и робототехника.* GPTA – робот может вести себя так, как ведет себя человек, набивая текст на клавиатуре, когда философствует.

GPTA не опровергает иные точки зрения на искусственный интеллект, в частности, на робототехнику. Просто утверждается, что GPTA не является интеллектуальной машиной. Подобная робототехническая ситуация номологически возможна, т.е. не противоречит науке, когда машина Блока, помимо диалога, вычисляет траектории элементарных частиц при нажатии на клавиатуру манипуляторами робота. Однако на самом деле этот GPTA не философствует. То, что он делает, – это вычисляет траектории элементарных частиц, чтобы подражать человеку, занимающемуся философией. Манипуляция роботом дескрипциями человеческих размышлений само по себе не является размышлением. Ни в



классическом, ни в общем, ни в сильном, ни в глобальном, ни еще каком-то ином смысле трактовки понятия искусственного интеллекта.

Таким образом, лингво-бихейвиористская способность искусственной системы продуцировать “осмысленные” ответы не является логически необходимой для приписывания интеллекта ГРТА-системе. В ГРТА нет ни ума, ни глупости. Она включает совокупность подстрок, каждая из которых бессмысленна. За исключением одного смысла – когда-то кем-то этот кусок был осмыслен как часть некоторой целостной осмысленной конструкции. Порядок этой совокупности не имеет смысла. За исключением одного смысла: когда-то кем-то этот кусок текста был осмыслен порядком следования фрагментов текста. В ответе как в статистически обусловленном результате генерации нет никакого смысла. Осмысленность ГРТА приписывается исключительно воображением наблюдателя.

“Мышление” машины – суть мышление экспертов (программистов, в первую очередь). ГРТА ни актуально, ни диспозиционально мыслить не может. Диспозиционально мыслит разработчик информационных средств ГРТА. Актуально мыслит пользователь ГРТА, приписывающий смысл продуктам ИТ.

С нашей стороны добавим еще один аргумент, который назовем *аргумент предикат предиката*. Он обусловлен психофункционалистским нюансом строения машины Блока. В аргументе Чизхолма-Гича (Block, 1981, pp. 11-12) демонстрируется идея системной обусловленности психологических феноменов и невозможности целочисленного кодирования диспозиций “вход – выход” из-за того, что представление некоторого феномена динамически расширяет состав связанных с ним феноменов. Поэтому требование построения динамической системы классификации и кодирования психических феноменов вряд ли сегодня можно прямолинейно реализовать, нужен язык второго порядка, определенный как “предикат предиката”. Однако второпорядковое понимание логического языка идет врозь с теорией и практикой современной вычислительной математики и информатики.

Персонофункционализм на порядок сложнее психофункционализма, если продолжить блоковское развитие формулировок “функционализма”. Дело в том, что “боль” – это весьма отчетливый феномен сознания. Его в самом деле можно обозначить, формализовать, указать диспозиции, выделить типовые и индивидуальные корреляты мозговой активности. А как быть с онтологическими параметрами личности – “смыслом”, “самостью”, “ценностью”, “творчеством” и пр.? По всей видимости, претензии ГРТА на персонофункционализм, т.е. на замену цифровыми помощниками естественных личностей, станут правдоподобными тогда, когда ГРТА воплотит проект включения ментального в машинно-функциональное, чтобы личности, и естественные, и искусственные, обживали “ментальную третью империю” Г. Фреге (Фреге, 2000, с. 471) или, скажем, “третий мир” К. Поппера, вместе со осмысленными высказываниями, истинными теоремами, оптимальными алгоритмами.



ЗАКЛЮЧЕНИЕ

Методологические исследования GPT-ассистентов, т.е. принципы построения и использования систем т.н. “генеративного ИИ», следует изучать на основе теста Блока как тьюрингового теста, реализуемого в условиях глобальных информационных ресурсов Интернета посредством нейрокомпьютерной генерации осмысленных подстрок как фрагментов осмысленного текста. Обобщенные функции машины Блока инварианты общим принципам работы генеративного ИИ. Архитектура и принципы работы машины Блока убеждают в том, что она никоим образом не объясняет принципы функционирования естественного интеллекта. Поэтому системы генеративного ИИ не способны претендовать на роль системы “сильного ИИ”.

Более того, понятие “генеративного ИИ” невозможно подвести под понятие “система искусственного интеллекта”, потому что принципиально элиминирует из системы управления способами представления знаний эвристики как правила формирования способов функционирования компьютерной системы ИИ. Понятие “генеративный ИИ”, сформулированный на основе GPT-систем представляется поспешным и нелегитимным. Введение этого термина на уровне государственных стандартов не прошло предварительных серьезных исследований и научных дискуссий.

Произведения GPT-ассистентов не критически интерпретируются как высоко осмысленные, разумные, креативные. При этом неясным остается юридическая правомерность коммерческого присваивания фирмами-разработчиками генеративных систем знаний различных форм человеческой культуры, переработка их в обезличенном формате с потерей и трансформацией смыслов.

Интересно было бы изучить причины, почему нелегитимные системы ИИ неожиданно получили статус легальных систем ИИ (Перечень поручений Президента РФ, 2023). От тех, кто критически оценивает стратегии развития ИИ и перспективы электронной культуры многое зависит, чтобы “генеративный искусственный интеллект” не послужил причиной глобального дегенеративного естественного интеллекта.

СПИСОК ЛИТЕРАТУРЫ

- Алексеев, А. Ю. (2008). Трудности и проблемы проекта искусственной личности. *Искусственные общества*, 1(3). <https://artsoc.jes.su/s207751800000077-3-1/>
- Алексеев, А. Ю. (2013). *Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты*. ИИнтелЛЛ.
- Алексеев, А. Ю. (2016). *Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга*. [Дис. ... д-ра филос. наук]. Моск. гос. ун-т им. М. В. Ломоносова.
- Алексеев, А. Ю. (2019). Общефункционалистский концепт искусственной потребности как основа общего искусственного интеллекта. *Философские науки*, 62(11), 111-124. <https://doi.org/10.30727/0235-1188-2019-62-11-111-124>



- Алексеев, А. Ю. (2020). Мультиагентные категории робофилософии: робочность и робозомби. *Искусственные общества*, 15(2). <https://doi.org/10.18254/S207751800009761-6>
- Алексеев, А. Ю., & Пожарев Т. (2020). Принципы креативной робототехники. *Технологос*, 4, 5-17. <https://doi.org/10.15593/perm.kipf/2020.4.01>
- Лекторский, В. А., Алексеев, А. Ю., Винник, Д. В., Гарбук, С. В., Черногор, Н. Н. (2022). *Методика оценки доверия к «искусственному интеллекту»*. Президиум РАН.
- Перечень поручений Президента РФ по итогам конференции «Путешествие в мир искусственного интеллекта», состоявшейся 23–24 ноября 2022 года. (2023). <http://www.kremlin.ru/acts/assignments/orders/70418>
- Фреге, Г. (2000). *Логика и логическая семантика*. Аспект Пресс.
- A.L.L. (1843). Sketch of the Analytical Engine invented by Charles Babbage, Esq. By L.F. Menabrea of Turin, Officer of the Military Engineers. In R. Taylor (Ed.), *Scientific Memoirs, selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals* (pp. 666 – 731). Richard and John E. Taylor.
- Bender, E., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* 🦜. In *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Block, N. (1981). Psychologism and Behaviorism. *The Philosophical Review*, 90(1), 5-43.
- Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, 9, 261-325.
- Bringsjord, S., Bello, P. & Ferrucci, D. (2001). Creativity, the Turing Test, and the (Better) Lovelace Test. *Minds and Machines*, 11, 3–27. <https://doi.org/10.1023/A:1011206622741>
- Chisholm, R. M. (1957). *Perceiving: A Philosophical Study*. Cornell University Press, Ithaca.
- Cole, D. (2023). The Chinese Room Argument. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/archives/sum2023/entries/chinese-room/>
- Frege, G. (2000). *Logic and logical semantics*. Aspect Press
- Geach, P. (1957). *Mental Acts. Their Content and their Objects*. Routledge & Kegan Paul.
- Levin, J. (2023). Functionalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/archives/sum2023/entries/functionalism>
- Miller, G.A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. Henry Holt & Co.
- Ned Block (2024, 22 February). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Ned_Block&oldid=1209557312



- Putnam, H. (1960). Minds and Machines. In S. Hook (Ed.), *Dimensions of Mind*. New York University Press. Reprinted in *Mind, Language, and Reality* (pp. 362 – 385). Cambridge University Press.
- Putnam, H. (1997). Functionalism: Cognitive Science or Science Fiction?. In D. M. Johnson and C. E. Erneling (Eds.), *The Future of the Cognitive Revolution* (pp. 32–44). Oxford University Press.
- Riedl, M. O. (2014). The Lovelace 2.0 Test of Artificial Intelligence and Creativity. *arXiv*. <https://doi.org/10.48550/arXiv.1410.6142>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/lix.236.433>

REFERENCES

- A.L.L. (1843). Sketch of the Analytical Engine invented by Charles Babbage, Esq. By L.F. Menabrea of Turin, Officer of the Military Engineers. In R. Taylor (Ed.), *Scientific Memoirs, selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals* (pp. 666 – 731). Richard and John E. Taylor.
- Alekseev, A. Yu. (2008). Difficulties and Problems of the Project of Artificial Personality. *Artificial Societies*, 1(3). <https://artsoc.jes.su/s2077518000077-3-1/>
- Alekseev, A. Yu. (2013). *Comprehensive Turing test: Philosophical and Methodological and Sociocultural Aspects*. AIntell.
- Alekseev, A. Yu. (2016). *Philosophy of artificial intelligence: the conceptual status of a Comprehensive Turing test*. [Dr. Phil. Dis.]. Moscow State University named after M. V. Lomonosov.
- Alekseev, A. Yu. (2019). The General Functionalist Concept of Artificial Need as the Basis of General Artificial Intelligence. *Philosophical sciences*, 62(11), 111-124. <https://doi.org/10.30727/0235-1188-2019-62-11-111-124>
- Alekseev, A. Yu. (2020). Multigenate Categories of Robofilosophy: Robolic and Robozobi. *Artificial societies*, 15(2). <https://doi.org/10.18254/s207751800009761-6>
- Alekseev, A. Yu., & Pozharev, T. (2020). The Principles of Creative Robotics. *Technologos*, 4, 5-17. <https://doi.org/10.15593/perm.kipf/2020.4.01>
- Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* 🦜. In *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, 9, 261-325.
- Block, N. (1981). Psychologism and Behaviorism. *The Philosophical Review*, 90(1), 5-43.
- Bringsjord, S., Bello, P. & Ferrucci, D. (2001). Creativity, the Turing Test, and the (Better) Lovelace Test. *Minds and Machines* 11, 3–27. <https://doi.org/10.1023/A:1011206622741>



- Chisholm R. M. (1957). *Perseving: A Philosophical Study*. Cornell University Press.
- Cole, D. (2023). The Chinese Room Argument. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*
<https://plato.stanford.edu/archives/sum2023/entries/chinese-room/>
- Frege, G. (2000). *Logic and Logical Semantics*. Aspect Press.
- Geach, P. (1957). *Mental Acts. Their Content and their Objects*. Routledge & Kegan Paul.
- Lectorsky, V.A., Alekseev, A. Yu., Vinnik, D.V., Garbuk, S.V., & Chernogor, N. N. (2022). *Methodology for Assessing Trust in “Artificial Intelligence”*. Presidium of the Russian Academy of Sciences.
- Levin, J. (2023). Functionalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*
<https://plato.stanford.edu/archives/sum2023/entries/functionalism>
- Miller, G.A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. Henry Holt & Co.
- Ned Block (2024, 22 February). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Ned_Block&oldid=1209557312
- Putnam, H. (1960). Minds and Machines. In S. Hook (Ed.), *Dimensions of Mind*. New York University Press. Reprinted in *Mind, Language, and Reality* (pp. 362 – 385). Cambridge University Press.
- Putnam, H. (1997). Functionalism: Cognitive Science or Science Fiction?. In D. M. Johnson and C. E. Erneling (Eds.), *The Future of the Cognitive Revolution* (pp. 32–44). Oxford University Press.
- Riedl, M. O. (2014). The Lovelace 2.0 Test of Artificial Intelligence and Creativity. *arXiv*. <https://doi.org/10.48550/arXiv.1410.6142>
- The List of Instructions of the President of the Russian Federation Following the Results of the Conference “Journey to the World of Artificial Intelligence”, held on November 23–24, 2022. (2023)
<http://www.kremlin.ru/acts/assignments/orders/70418>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.
<https://doi.org/10.1093/mind/lix.236.433>

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Андрей Юрьевич Алексеев, aa65@list.ru,
ORCID 0000-0003-2230-311X

Andrey Yurievich Alekseev, aa65@list.ru,
ORCID 0000-0003-2230-311X

Екатерина Алексеевна Алексеева,
eaalekseeva@gagn.ru
ORCID 0000-0002-0006-5942

Ekaterina Alekseeva Alekseeva,
eaalekseeva@gagn.ru
ORCID 0000-0002-0006-5942

Статья поступила 20 февраля 2024
одобрена после рецензирования 11 апреля 2024
принята к публикации 3 июня 2024

Received: 20 February 2024
Revised: 11 April 2024
Accepted: 3 June 2024