



Artificial societies. 2013-2020

ISSN 2077-5180

URL - <http://artsoc.jes.su>

All right reserved

Issue 4 Volume 14. 2019

The Future of Artificial Intelligence: Turing or Post-Turing Methodology?

A. Alekseev

*Moscow State University
Russian Federation, Moscow*

A. Efimov

*Sberbank
Russian Federation, Moscow*

V. Finn

*Institute of Informatics Problems of the Federal Research Center "Informatics and Management" of
the Russian Academy of Sciences
Russian Federation, Moscow*

Abstract

Can it be argued that the Turing test has already been passed and the development of intelligent systems is carried out within the framework of some post-turing methodology, where the severity of the problem of distinguishing between natural and artificial intelligence has been removed? It is also possible that the approach to AI as a computer simulation of natural intelligence leads us into a dead end and therefore the future of AI should be associated with the idea of strengthening natural intelligence?

Keywords list (en): artificial intelligence, Turing test, Turing private test, Turing comprehensive test, intelligent systems

Date of publication: 19.12.2019

Citation link:

Alekseev A., Efimov A., Finn V. The Future of Artificial Intelligence: Turing or Post-Turing Methodology? // Artificial societies. 2019. V. 14. Issue 4 [Electronic resource]. Access for

1 **Ефимов А.Р.** Мне бы хотелось начать нашу дискуссию с двух утверждений. Во-первых, антропоцентрический вектор всех исследований в области искусственного интеллекта за все годы развития ИИ предзадан Аланом Тьюрингом в 1950 году [Turing, 1950]. Ни одно из фундаментальных открытий в области компьютерных наук не предопределило направление всех исследований и разработок больше, чем сформулированная Тьюрингом идея о сравнении рассуждений человека и машины. Позже эта идея была названа тестом Тьюринга. Несколько раз появлялись сообщения о создании диалоговых систем, которые могут выдавать себя за человека в течение того времени, которое сам Тьюринг считал достаточным для распознавания сущности скрытых от наблюдателя собеседников. Эти сообщения провоцировали мнения в обществе о том, что тест Тьюринга уже пройден или, по крайней мере, мы очень близки к его прохождению. Но в рамках нашей дискуссии я хотел бы рассмотреть гипотезу о том, что тест Тьюринга в своем изначальном виде вовсе не являлся барьером, который надо было взять. Современным диалоговым системам, возможно, и не надо его проходить, чтобы доказать «искусственный интеллект создан и существует» по причине того, что тест Тьюринга перестал быть релевантным и человек-собеседник больше не проецирует антропный принцип на своего искусственного собеседника. Иными словами, искусственному интеллекту более не надо выдавать себя за умелого собеседника-человека, который настолько похож на человека, что мы не можем распознать искусственность. Искусственный собеседник с самого начала дает человеку опознать свою «искусственность», но при этом диалог с таким собеседником может оказаться полезнее для человека, чем диалог с обычным человеком. Тест Тьюринга оказался своего рода «*линией Мажино*» - он не был взят прямым штурмом, но умело обойден всем ходом развития компьютерной науки и технологии. Искусственный интеллект стал не антропоморфным инструментом решения проблем для человека, но стал важной частью самого человека. Именно в этом направлении идет развитие всех коммерческих промышленных голосовых и виртуальных ассистентов (Siri, Google Now, Cortana, Алиса и других).

2 Во-вторых, не является ли современный искусственный интеллект полностью «искусственным» и уж точно никаким не «интеллектом». Может быть, мы слишком сильный акцент ставим на «искусственном» и слишком мало внимания уделяем «интеллекту»? ИИ, возможно, все также «туп», каким был полвека назад, когда Джозеф Вейценбаум разработал первую диалоговую компьютерную программу ELIZA. По его словам, когда он застал своего секретаря за общением с ELIZA посредством телетайпа (тогда общались с компьютером с помощью телетайпа), то секретарь обратилась к профессору с просьбой: «Выйдите, пожалуйста, у меня слишком интимный диалог с моим психиатром» (Weizenbaum J., 1976). Сам Дж.Вейценбаум считал ELIZA примитивной программой. Однако ни один современный интеллектуальный ассистент до сих пор не вышел за принципы ее функционирования. Причина такого медленного прогресса в том, что человек обычно ведет диалог в тесной связи с контекстом общения, тогда как для интеллектуальных помощников это пока недостижимо.

3 Давайте сначала поймем, где мы сейчас находимся в отношении искусственного интеллекта. Можем ли хоть как-то определить это поле исследований, и назвать это поле «искусственным интеллектом». Джон Маккарти дал четкое определение: «искусственный интеллект – инженерно-математическая дисциплина, занимающаяся созданием программ и устройств, имитирующих когнитивные (интеллектуальные) функции человека, включающие в том числе анализ данных и принятие решений». Да, я уверен, что большинство помнит его. Но, может быть, это определение уже давно потеряло свою актуальность, как и тест Тьюринга. Андрей Юрьевич, может быть Вы начнете? Что Вы думаете по поводу определения?

4 **Алексеев А.Ю.** Слово сочетание «тест Тьюринга» сейчас у всех на слуху. Предполагается, что этот тест сформулировал Алан Тьюринг в статье «Вычислительные машины и интеллект» 1950 г. Однако это не так. В этой статье никакого теста нет. Слово «тест» в ней не употребляется в смысле выявления интеллекта у испытуемых. А. Тьюринг предлагает поиграть в игру в имитацию, участниками которой являются судья (наблюдатель, исследователь), мужчина, женщина и компьютер. Цель судьи – определить пол игрока: с мужчиной или женщиной ведет он беседу? На первом этапе играют люди – мужчина и женщина. Судья их не видит и общается с ними письменными сообщениями телетайпа, то есть тем, что сегодня называется электронной почтой, чатом, ботом, но не скайпом и тем более не мультимедийными средствами передачи запахов и рукопожатий, – судья не видит собеседника. Женщина говорит только правду, мужчина может вводить судью в заблуждение, прикидываясь женщиной. Конечно, при таких условиях судья только наугад способен предсказать правильный пол: «50 на 50, и зачем это надо?» – примерно так отреагировала Софья Александровна Яновская, автор предисловия к книге «Могут ли машины мыслить?» (1960 г.), создатель советской школы философии математики и организатор кафедры математической логики на механико-математическом факультете МГУ.

5 Однако логическая корректность сценария не волнует автора статьи. Ведь главное – продемонстрировать способность имитации компьютером человеческого диалога. Важно то, насколько часто судья будет ошибаться при общении с человеком и компьютером. Условно введенная вероятность ошибки – 30 процентов неправильного определения пола при пяти минутах «общения», хотя и теряет смысл в контексте некорректных начальных условий, тем не менее, не существенна для обозначения принципа компьютерной имитации интеллектуальных способностей, проявляемых в условиях повседневного общения людей. Для этого на второй стадии игры в имитацию роль женщины играет компьютер (здесь и возникают отмеченные выше казусы с вероятностями). А на третьей стадии компьютер уже «захватывает» роль мужчины, пол игрока становится несущественным, и, собственно, возникает возможность сформулировать то, что принято считать каноническим тестом Тьюринга, точнее, его двумя формами. Имеются две разновидности канонического теста Тьюринга. Они сегодня общеприняты. Первая имеет дистрибутивный, разделительный формат, а вторая – атрибутивный формат, приписывающий человечность или компьютерность тестируемой системе.

6 Дистрибутивная форма. В игру имитации интеллекта играют и люди и компьютеры. Задача судьи – определить, кто есть кто? Кто из собеседников – компьютер, а кто – человек? Эта форма сейчас используется в Лойбнеровских состязаниях. Об этих «тьюринговых скачках», по всей видимости, расскажет Игорь Пивоваров.

7 Атрибутивная форма. В игру в имитацию интеллекта играет неизвестный субъект. Задача судьи — приписать собеседнику либо «человечность» либо «машинность». Это взаимосвязанные формы: определить – значит разграничить и наоборот. И та и другая формы значимы. Например, они непосредственно характеризуют информационную безопасность. Для судьи важно не различие компьютерного/человеческого, а выявление факта общения именно с человеком. С кем я пойду сегодня на свидание? Кто просит меня переслать электронные деньги? Кто мне навязывает политический имидж президента? Это – автономная программа или реальный человек? Какие фрагменты «человека-компьютера» можно запрограммировать и кто на это способен?

8 Как я отметил выше, сам А. Тьюринг в этой статье не употреблял слово «тест» в смысле тестирования компьютера на интеллект. И это послужило поводом для ехидных замечаний: в мире когнитивной философии А. Тьюринг стал известен за то, что, собственно, никогда и не предлагал – он не сформулировал «тест Тьюринга» (Дж. Лассаж, 1996). Однако

это не совсем так. На самом деле, А. Тьюринг в 1952 году в репортаже на студии БиБиСи говорит о своем мысленном эксперименте не как об игре в имитацию, но как о тесте. Правда, в этой беседе, указав на сложность задачи моделирования мозга и интересубъективных отношений, А. Тьюринг отодвинул сроки реализации такого теста еще на пятьдесят лет вперед [9, Р. 126]. То есть, по его новому предсказанию, не в 2000 году, а в 2052 году следовало бы ожидать появление «компьютерного собеседника по переписке от рождения до смерти». Если бы прогноз сдвигался и далее с такой прогрессией динамики «знаний о незнании» (ведь в 1950 году А.Тьюринг не знал о сложности задач, которую оценил спустя два года), то в этом, 2019 году, следовало бы ожидать прохождения теста Тьюринга не ранее 3872 года. Это было бы к великому сожалению для устроителей Лойбнеровских состязаний, стремящихся буквально следовать тонкостям тьюринговой идеи игры в имитацию. Но, наверное, к радости Игоря Пивоварова, потому что он свои конференции по тесту Тьюринга мог бы продолжать до указанного года и жить вечно.

9 Конечно, догматичному толкованию игры в имитацию не следуют нормальные исследователи теста Тьюринга. Для них тест служит концептуальным каркасом искусственного интеллекта, является генератором новых идей, различных интерпретаций и толкований. Исследуются самые разные параметры теста: предмет тестирования – интеллект, сознание, переживание и пр.; длина теста – интервал времени, в течение которого осуществляется тестирование; содержание тестовых вопросов; стена Тьюринга – параметры скрытости игроков от судьи; особенности игроков и судей и многое др. На разнообразие трактовок влияет неопределённость идеи игры в имитацию, «чехарда» с ролями мужчины, женщины, машины и пр.

10 Реакция на статью возникла, впрочем, не сразу: 15 - 20 лет дискуссии не выходили за пределы журнала «Mind». Причем вряд ли эти дискуссии можно назвать серьезными. Так, Кейт Гандерсон в статье 1964 г. предлагает поиграть в игру «в наступалки»: судья просовывает ногу в дырку стены комнаты и на его ногу начинают наступать в различном порядке мужчина и женщина. Далее вместо женщины начинает работать машина, которая посредством видеодатчиков запускает механизм падения камней на просунутую ногу. Далее эта машина начинает играть роль мужчины. Получается вопрос «Могут ли камни подражать?». И что из этого следует? Другие «исследователи» вместо ленты Тьюринга предлагают использовать туалетную бумагу и поднимают вопрос: какой длины должна быть бумага, чтобы возникла иллюзия имитации интеллекта? А Ричард Пёртилл в 1971 г., выступая как «продвинутый» программист, но не как философ, выказывает готовность съесть электронную библиотеку у того, кто спроектирует компьютер, успешно играющий в игру в имитацию. Игра в имитацию – это из сферы научной фантастики, считает он. Однако в этом же году К. Колби в работе «Искусственная паранойя» (1971 г.) представил описание обильно профинансированной реализации игры в имитацию компьютером ответов пациентов психиатрической лечебницы. Врачи–судьи затруднялись ответить, от людей или компьютеров поступают ответы. Тест Колби называют параноидальным тестом Тьюринга.

11 Так вот. Первые пятнадцать лет никто не говорил о тесте Тьюринга. Говорилось именно об игре в имитацию. Статьи, в которых начинает фигурировать «тест Тьюринга», датируются концом 1960-х годов. Некоторые исследователи занялись плагиатом тьюринговой идеи. Так, Джон Сёрль в 1980 году в своем знаменитом мысленном эксперименте «Китайская комната» почти досконально воспроизвел сценарий Тьюринга, лишь немного его изменив: роль человека и компьютера играет сам Дж. Сёрль, снабженный компьютером с базой данных-иероглифов, а роль судьи попеременно играют китаец и англичанин. Ну не его этот сюжет: Дж. Сёрль на самом деле забрел в комнату Тьюринга, но хоть как-то для приличия надо бы было обозначить ссылку на сюжет тридцатилетней давности.

12 Многие исследователи, однако, отнеслись более честно и серьезно к игре в имитацию. Например, Нэд Блок в 1981 году предлагает т.н. «новый тест Тьюринга», в котором обращает внимание на важность смысловых параметров диалога, а не только на синтаксис и семантику сообщений. Также он формулирует отличие интеллектуальной системы от системы искусственного интеллекта, предвосхищает роль Интернета как глобальной интеллектуальной компьютерной системы, в которой, однако - доказывает он - нет ни толики искусственного интеллекта, формулирует концепцию психофункционализма – одного из продуктивных направлений моделирования когнитивных систем и пр.

13 В Советском Союзе отношение к идее А. Тьюринга с самого начала было весьма почтительным. Во-первых, следует учесть тираж книги «Могут ли машины мыслить?» – 50 тысяч экземпляров! Причем книга мгновенно распространилась по всей стране, по словам Б.В.Бирюкова, редактора книги. Сразу повсеместно разгорелись жаркие дискуссии, возникли научно-фантастические фильмы и повести про искусственный интеллект. В 1963 г. академик А.Н. Колмогоров пишет статью «Автоматы и жизнь», где утверждает теоретическую возможность воспроизводства автоматами всех видов человеческой активности и не только интеллектуальных, но и эмоциональных. На это Б. Бялык отвечает статьей «Товарищи, вы это серьезно?». В ответ академик С. Л. Соболев пишет работу «Да, это вполне серьезно!». Завязалась серьезная полемика, в рамках которой обсуждались вопросы возможности кибернетического определения таких понятий, как воля, мышление, эмоции и др.; противоречит либо соответствует материалистической диалектике концепция «думающих» машин; как понимать «машину» в контексте исторического подхода, ведь машина – это продукт общественно-трудовой деятельности человека, машина не трудится, трудится человек посредством машины и пр. Возникли этические проблемы контроля над «интеллектуальными» машинами. Вскоре в нашей стране споры по поводу мыслящих машин, однако, утихли. Как, впрочем, были свернуты интереснейшие проекты советских ЭВМ, таких как трехзначная машина «Сетунь» (разработка МГУ). Примерно в эти же годы протоколы вычислительных сетей, разработанные в отечественных НИИ, оказались в ARPANET под новым названием «IP». Произошел ряд других событий, непонятных с позиции стратегии национального научно-технического прогресса.

14 За рубежом интерес к тесту Тьюринга ярко вспыхнул в 1990-е годы. Появились крупные статьи. В аналитической философии сознания прочные позиции занял компьютерный функционализм. В 1991 году продавец светящихся полов для дискотек Хьюго Лойбнер (Hugh Loebner) открыл ежегодные тьюринговые игры, впервые предложив 100 тысяч долларов за программу, которая пройдет тест Тьюринга (точнее, игру в имитацию). В общем, тьюринговая гонка началась.

15 Сегодня насчитываются сотни крупных статей, монографий, диссертационных исследований и специальных работ, посвященных тесту Тьюринга. Их можно обозначит как «частные тесты Тьюринга». В литературе выделяется определенное количество крупных версий теста Тьюринга. Условно говоря, имеется двадцать одна версия. Эти версии обозначены как совершенные частные тесты Тьюринга, так как в них четко просматриваются базовые функции, предложенные в оригинальном тесте:

- 16 1. интеррогативная функция, характеризующая содержание и форму вопросов частного теста, которые задает тьюринговый судья (interrogator, наблюдатель) при изучении х-системы на предмет ее у-способностей;
2. дефинитная функция, обеспечивающая компьютерное определение когнитивного феномена;
3. конструкторская функция, раскрывающая принципы работы компьютера, способного реализовать когнитивный феномен;

4. критическая функция, отражающая суть полемики по поводу возможности компьютерной реализации когнитивного феномена;
5. конститутивная функция, позиционирующая отношение судьи (наблюдателя) к когнитивным аспектам компьютерной реальности.

17 Помимо совершенных версий, имеется большое количество несовершенных. В них акцентируется внимание на интересной, но незначительной детали тьюрингового тестирования. Таких несовершенных версий много, более сотни зарегистрированных. Кратко перечислю совершенные частные тесты, которые, на мой взгляд, составляют ядро комплексного теста Тьюринга. Подробно они рассмотрены в книге [7] и концентрировано представлены на страницах журнала «Искусственные общества» в 2011 году:

- 18 1. Оригинальный тест. Отвечает на вопрос: может ли компьютер мыслить? (А. Тьюринг, 1950). Изучает возможности компьютерной имитации осмысленного диалога. Применяет концепцию универсальной УЦВМ и параллельной УЦВМ – так называемой «нейронной сети Тьюринга».
2. Тест на «здравый смысл». Включает тест на глубинный «здравый смысл»: может ли компьютер разумно рассуждать? (Дж. Маккарти, 1984) и тест на поверхностный «здравый смысл»: может ли компьютер банально рассуждать? (Д. Деннет, 1984). Экспертные системы не способны отвечать ни на глупые вопросы, ни на вопросы, требующие глубокого понимания предмета.
3. Тест «Китайская нация»: возможна ли компьютерная реализация общественного сознания? (Н. Блок, 1978, 1981). Глобальный искусственный интеллект как сингулярная сущность невозможен, он возможен лишь как сочетание деятельности многочисленных разработчиков-людей, каждый из которых обладает уникальным естественным интеллектом.
4. Тест «Китайская комната» (тест Серля): может ли компьютер понимать? (Дж. Серль, 1980 г.). Компьютер понимать не может, более того, мешает процессу «понимания». Он не способен к интенциональности и является всего лишь формальной «тенью» активного осмысления предметной области, на что способен только человек в силу биологического прошлого.
5. Параноидальный тест: может ли компьютер разумно функционировать, как нормальный, психически здоровый человек? (К. Колби, 1971). Нет, не может. Экспертные системы способны имитировать только психические отклонения, паранойю. Параноидальные представления – это идеи программиста, границы которых непреодолимы.
6. Субкогнитивный тест (тест Френча): может ли компьютер обладать «подсознанием»? (Р. Френч, 1990). Компьютер должен «жить» среди людей, тогда его «база знаний» будет наполняться ассоциациями повседневного опыта. При таких условиях он будет проходить тест.
7. Эмоциональный тест: может ли компьютер любить (испытывать эмоции)? (А. Сломан, 1998 г., Дж. Маккарти, 2001). Компьютер способен «любить», но лишь по-функционалистски, без эссенциальных притязаний.
8. Инвертированный тест: может ли компьютер приписывать ментальность х-системам? (С. Ватт, 1996 г.). Компьютер не может приписывать ментальное, у него отсутствует способность к «воображению». Как только это обнаруживается, значит, мы имеем дело с человеком, но не с компьютером.
9. Гендерный тест: может ли женщина (мужчина) мыслить? (Ю. Генова, 1994). Изучает вопросы гендерной детерминации интеллектуальной деятельности. Включает тест Хопкинса (Р. Хопкинс, 1998 г.): может ли компьютерным образом организованная система испытывать сексуальное влечение и наслаждение? Поднимает вопросы

алгоритма работы «секс-машины».

10. Креативный тест (тест Лавлейс): может ли компьютер творить? (С. Брингсйорд, 2001). Компьютер не может творить. Да, созданы искусные программы генерации высокохудожественных литературных, изобразительных и музыкальных произведений. Однако они «творят» в контексте известного.
11. Тест Гёделя-Лукаса-Пенроуза: способен ли компьютер к самоописанию и самоорганизации? (Дж. Лукас, 1961; Р. Пенроуз, 1989). Компьютер не способен ни к самоописанию, ни к самоорганизации. Впрочем, человек тоже на это не способен.
12. Индуктивный тест (тест Мура) как достаточное логическое условие искусственного интеллекта. Это условие индуктивной аргументации в пользу возможности построения искусственного интеллекта. Оно достигается путем коллекционирования программ, успешно прошедших тест Тьюринга, и отбраковки программ, не прошедших этот тест (Дж. Мур, 1976).
13. Новый тест Тьюринга (тест Блока): может ли компьютер осмысленно отвечать на любой синтаксически корректный вопрос? (Н. Блок, 1980). Прохождение теста Тьюринга — необходимое логическое условие искусственного интеллекта. Условие достижимо, но для этого надо включить разработчика в состав компьютерной системы. Но в этом случае мы получим не «искусственный интеллект», но лишь формализованный интеллект разработчиков, т.е. то, что называется «интеллектуальной системой».
14. Кибериадный тест: может ли компьютер творить искусственные миры? (Дж. Баресси, 1987). Предполагается иерархия миров, в которых действуют компьютерные законы создания «себе подобных». Окрашивает некоторые религиозные вопросы компьютерными коннотатами.
15. Тотальный тест (тест Харнада): может ли компьютер стать неотличимым от человека? (С. Харнад, 2001). Допустим, можно создать компьютерный дубликат человека, то есть реализовать его точную макро- и микрофизическую копию. Однако будет ли дубликат мыслить и осознавать? В свете этого вопроса все современные достижения искусственного интеллекта – лишь игрушечные поделки. Перспективно применение этого теста в конвергентной методологии НБИКС. «Самый тотальный тест» преследует более радикальные креационистские цели и изучает вопросы прохождения компьютером пути социо-когни-био-эволюции и творения компьютером иных компьютерных миров (П. Швайзер, 1998).
16. Нейрокомпьютерный тест или «тест Тьюринга с мозгом» (Черчленды). Классический компьютер не может мыслить, нейрокомпьютер – может. Для этого требуется программа элиминации нейрофизиологических терминов относительно нейрокомпьютерных, как ранее были элиминированы термины народной психологии относительно нейрофизиологических терминов.
17. Тест искусственной личности: может ли компьютерная система стать личностью? Различаются проекты: искусственная личность как антропоморфный робот (Д. Деннет, 1994; Дж. Поллок, 1995) и как экспертная система, репрезентирующая надчеловеческие «знания» (А. Алексеев, В. Деев, 1994).
18. Тест философских зомби: можно ли имитировать сознательное существо посредством бессознательной системы? (Р. Кирк, 1971). Имеет много субтестов. Например, тест Чалмерса: могу ли я быть компьютерным зомби? (Д. Чалмерс, 1995). Для этого нужен компьютер, построенный в соответствии с «новой теорией информации». Что это за теория? Компьютерная реализация философских зомби – это всего лишь компьютерная экспликация художественных образов. Проект искусственной личности, включающий блок «псевдосознания», более правдоподобен (Дж. Маккарти, 1998).
19. Интроспективный тест: может ли компьютер обладать «иным» сознанием? Компьютерная система способна обладать психикой нечеловеческого рода (А. Клифтон,

2003). Более того, глобальный искусственный интеллект (у него-то явно «иное» сознание) обладает настоящим сознанием, а все люди – зомби, так как не владеют полнотой сознательного опыта (Д. Деннет, 2001). Поэтому целостность когнитивной жизни не в естественном биологическом прошлом человека, как считает Дж. Серль, но в искусственном технологическом будущем человечества, как это доказывает Д. Деннет.

20. Интерактивный тест (тест Финна, 2009): может ли компьютер работать с «идеями» как с бесконечно становящимися понятиями, т. е. как с философскими категориями? Тест востребован требованиями рационализации общественной жизни средствами интеллектуальных технологий и предназначен для изучения вопросов усиления естественного интеллекта. Для реализации теста применяется квазиаксиоматическая компьютерная система поиска закономерностей в фактах, генерации гипотез и верификации/фальсификации фактов. Об этом, по всей видимости, расскажет Виктор Константинович. Но, забегаая вперед перед его выступлением, отмечу, что оригинальный тест Тьюринга с его априори разработанной программой не нужен для технологии интеллектуальных систем. Конечно, требуется взаимодействие, интеракция между человеком и компьютером.
21. Комплексный тест: может ли компьютер ВСЁ? То есть *всё* то, на что способны частные тесты Тьюринга (А. Алексеев, 2006). Предполагается исследование возможности компьютерной реализации когнитивных феноменов самого широкого спектра. Для реализации теста предлагается машина Корсакова-Тьюринга, которая совмещает коннекционистскую и символьную парадигмы представления «знаний» интеллектуального компьютера.

19 Все эти тесты по отдельности интересны. Но в сочетании с другими тестами эффект получается, как говорят системотехники, эмерджентным. Например, возьмем тест № 2 (на «здравый смысл»). Интересно отметить, что эти тесты предложены в статьях, которые были опубликованы почти одновременно. Авторы их – крупнейшие когнитивные ученые – Дж. Маккарти и Д. Деннет. Первый вообще является основателем науки «искусственный интеллект». Они критикуют экспертные системы в самый разгар их завораживающих успехов. Однако имеются ли успехи, или это очередной технологический пиар? Что примерно сейчас вновь происходит с ИИ во все мире.

20 Предмет тестирования, казалось бы, одинаков – «common sense». Когда Д. Деннет задает экспертной новостной системе SYRUS вопрос, может ли Сайрус Вэнс (тогдашний генеральный секретарь при президенте Картере) завязывать шнурки, то компьютер не отвечает на этот пустяк, хотя досконально «знает» где, с кем, когда и по какому поводу был Сайрус Вэнс. Дж. Маккарти предлагает проверить медицинскую систему MYSIN с позиции высококлассного врача и обнаруживает, что настоящих-то знаний нет, пациент точно умрет, если доверится компьютеру.

21 К сожалению, нет возможности подробно останавливаться ни на этих тестах, ни на других тестах, ни на интересных эффектах их совместного применения. Но это уже – комплексный формат. То есть это и есть то, что можно обозначить как *комплексный тест Тьюринга*. Так вот, возвращаясь к идее нашей дискуссии. О какой тьюринговой и посттьюринговой методологии ИИ следует нам говорить? Если мы под тьюринговой методологией понимаем тот формат программирования и применения интеллектуальных систем, который подпадает под схему игры в имитацию, под оригинальный тест Тьюринга, то тогда посттьюринговая методология – это то, что задается комплексным тестом Тьюринга. Но что-то здесь не так. Надо послушать других спикеров.

проблематики ИИ начиная с 70-х годов (точнее, с 1975 года). Меня вовлек в эту деятельность, связанную с ИИ, создатель отечественного направления в области ИИ Дмитрий Александрович Поспелов. К сожалению, с 1998 года он парализован, и не принимает участие в исследовательской жизни. И более того, мне пришлось неоднократно общаться с профессором Джоном Маккарти. В частности, я бы хотел обратить внимание присутствующих на одно важное высказывание профессора Маккарти из статьи «Некоторые философские проблемы создания ИИ». Это работа 1969 года, она опубликована в сборнике проблем бионики в 1972 году. Джон Маккарти высказал такое соображение: «Работа над ИИ, особенно общим ИИ, пойдет гораздо успешнее, если будет уточнено само понятие интеллекта». Я думаю, что то возбуждение, которое возникло в последние годы в нашем обществе в связи с ИИ, не совсем оправдано. Понятие ИИ используется весьма расплывчато и мало кто понимает, каково академическое развитие этой области. Технологии, которые сейчас торжествуют, это ведь некоторый этап, причем весьма не решающий этап решения основных проблем ИИ.

23 Я неоднократно пытался определить ИИ. В частности, этому посвящено мое публичное выступление в газете «Независимая газета» 27 июня 2018 г. Называлось оно так «Тринадцать признаков ИИ». Если мы не представляем себе, что мы должны усиливать и имитировать, то мы становимся только рабами своих практических достижений. Ибо без теоретических оснований, без системы понятий, дальнейшее развитие технологий не будет успешным и разнообразным. Хотя определенные достижения, несомненно, имеют место, это, прежде всего нейронные сети и машинное обучение. Нейронные сети, однако, вызвали серьезную ошибку в ИИ, ибо отождествили ИИ как направление с машинным обучением. На мой взгляд, это не только не правильно, но и вредно.

24 Итак, вернемся к идее Дж.Маккарти об уточнении понятия ИИ. Что это такое? Во первых, ИИ – это не продукт, а некоторое направление научное и технологическое. Поэтому, когда говорят: «У меня работает ИИ», то это детский лепет, не серьезный разговор. И так с чего надо начать, чтобы определить ИИ как научное направление. Начать нужно с интеллектуальных способностей, которые мы способны имитировать или усиливать. В моем популярном изложении «Тринадцать признаков ИИ» я как раз и перечислил эти способности. Я начну их перечислять, потому что из их содержания следует отношение к тесту Тьюринга.

- 25 1. Способность естественного интеллекта – это выделение существенного в обозреваемых данных. Эту черту Карл Ясперс (выдающийся психиатр и философ) считал главной чертой естественного интеллекта.
2. Способность целеполагания, порождение последовательности «цель план действие».
3. Способность отбора знаний, релевантных цели рассуждения.
4. Способность рассуждения как извлечения следствий из имеющихся посылок в соответствии с целью, при этом важно, что рассуждение должно иметь амплиативные (термин Ч.С. Пирса) выводы, то есть выводы, отличные от дедукции, где из истины следует истина, но не извлекается новое знание из имеющихся посылок. Примеры амплиативных выводов: индукция, аналогия, абдукция (Пирс), получение гипотез с помощью объяснения.
5. Аргументация, принятие решений на основе сопоставления аргументов «за» и аргументов «против».
6. Рефлексия – умение оценивать свои знания и действия.
7. Способность к объяснению.
8. Одна из главных способностей для создания ИИ – это синтез познавательных процедур, решение задачи с помощью эвристики, которая применяет разные познавательные процедуры, не только индукцию и не только абдукцию, а их взаимодействие.

9. Обучение и использование памяти.
10. Рационализация неточных идей и превращение их в точные понятия.
11. Создание целостной картины изучаемого явления.
12. Познавательное любопытство, ответ на вопрос: «Что такое?»
13. Адаптация системы знаний в результате получения новых знаний и изменения ситуации.

26 Вот если мы имеем эти тринадцать способностей, то мы можем охарактеризовать теоретический интеллект. Заблуждением глубоко чтимого профессора Джона Маккарти было мнение, что надо изучать мышление ребенка и common sense интеллект. На мой взгляд, это тупиковый путь и он не дал позитивных результатов. Об этом Джон Маккарти писал в последних статьях, но я думаю, что это была ошибка. На самом деле мы должны исследовать теоретический интеллект и его приближения, и вот тут мы можем ответить на вопрос о роли теста Тьюринга. Дело в том, что из этих 13-ти способностей только некоторые могут быть реализованы в автоматическом режиме, остальные могут быть реализованы только в интерактивном режиме. Отсюда становится понятным, что тест Тьюринга может быть применен тогда и только тогда, когда мы имеем дело с имитацией тех способностей, которые могут быть реализованы в автоматическом режиме. Те способности, которые не могут быть реализованы в автоматическом режиме, не подлежат возможности сравнения с помощью теста Тьюринга, хотя тест Тьюринга, как некоторый критерий для сравнения с некоторыми системами имеет смысл.

27 Недавно мой коллега Михаил Забейало указал мне на наличие в интернете некоторой дискуссии по поводу того, в чем состоят точки роста в области ИИ. Их три. Это предлагают ученые из IBM, Стэнфорда и другие). Точки роста следующие.

- 28 1. Создание партнерских человеко-машинных систем. Здесь обязательно участие человека в процессе решения проблем с помощью компьютерных систем. Отсюда следует то, что как раз из этих 13-ти способностей, о которых мы говорили, некоторые могут быть решены только в интерактивном режиме. Следовательно, партнерство с человеком обязательно. Поэтому спекуляции на тему «суперразума», который когда-то будет мешать человечеству, не основательны. В качестве примера можно привести работы Рея Курцвейла, одного из таких спекулянтов.
2. Включение машинного обучения в общую теорию рассуждений, а не сведение решения задач к машинному обучению как таковому. Это очень серьезная проблема и связана она с тем, что здесь нужно реализовать одну из главнейших способностей: синтез познавательных процедур (взаимодействие индукции, аналогии, абдукции и дедукции)
3. Интеллектуальные роботы. Мы понимаем, что развитие робототехники может быть развито в разных направлениях (роботы – манипуляторы, роботы, решающие отдельные задачи) разумеется - это важное технологическое дело, оно будет развиваться. Но интеллектуальный робот, с помощью которого мы можем решать антропоморфные задачи, это то, что основано на интеллектуальных системах.

29 И вот чего не поняли многие западные исследователи, они не поняли одну вещь, и это ошибка профессора Маккарти, так как он не дал вектор. Дело в том, что ИИ как направление имеет три главных продукта, и только три.

- 30 • Первый продукт – это системы ИИ. Это применение отдельных методов и соответствующих программ для решения частных задач. В частности, нейронные сети.
- Второй и главный продукт – это интеллектуальные системы. Термин, который был введен нами. Но в дальнейшем стал использоваться расплывчато.

- Третий продукт – это интеллектуальный робот.

31 Остановимся на главном продукте – интеллектуальных системах. Что это такое? Интеллектуальной системой мы называем систему, которая имеет следующую архитектуру:

- 32 • Открытая и пополняемая база фактов.
- Открытая и пополняемая база знаний.
- Решатель задач и его модули.
- Рассуждатель, реализующий синтез познавательных процедур.
- Вычислитель, реализующий вычислительные процедуры.
- Синтезатор, который выбирает стратегии и координирует вычисления.

4. Комфортный интерфейс.

33 Но если мы имеем эту архитектуру, мы еще не имеем интеллектуальную систему (ИС). ИС с такой архитектурой будет только в том случае, если решатель задач реализует имитацию и усиление тех 13-ти способностей, о которых мы говорили выше. Если это происходит, то тогда эта система является ИС. Вот это главный продукт и естественно он – это реализация теоретического интеллекта и его развитие для различных решений классов задач. Это дело будущего и настоящего.

34 И наконец, где применяется ИС? Это те области знания и практики, где знания слабо формализованы, а данные сложно структурированы. Следовательно, это, прежде всего, науки о жизни и науки о социальном поведении. Там, где есть математический аппарат ИИ, делать нечего. Он востребован в широкой области наук о жизни и наук о социальном поведении. Это социология, медицина, управление, вот это сфера действия ИИ. Это основные соображения, которые я хотел сказать. Могу посоветовать всем интересующимся проблемами ИИ две книги. Это книга «Будущее искусственного интеллекта», 1991 года, под редакцией Дмитрия Александровича Поспелова, где высказываются наши с ним предположения о будущем ИИ [Будущее искусственного интеллекта, 1991]. Могу сказать, что по большей части мы не предугадали того, что произойдет, но некоторые мысли здесь были высказаны правильно. Я полагаю, что она интересна. И моя собственная книга «Искусственный интеллект», где я подробно обсуждаю все 13 тезисов. И рассуждаю об ИС в том смысле, в котором они были мной определены.

35 Коротко, я бы хотел прокомментировать то, что имеется возможность сформулировать определение ИИ. Это есть научное направление, которое занимается имитацией и усилением познавательной деятельности и рационального поведения человека. Это определение опирается на следующее понимание познавательной деятельности. Познавательная деятельность состоит из анализа данных, предсказания и объяснения. Главная проблематика ИИ состоит в имитации и усилении интеллектуального процесса. Интеллектуальный процесс состоит из мыслительного процесса и его взаимодействия с познавательным процессом. Важное соображение в том, что познавательный процесс не сводится к деятельности мозга. Думать, что если мы моделируем мозг, мы решаем проблемы познания ошибочно. Замечательная книга Карла Поппера важна для всех людей занимающихся ИИ: «Знание и психофизическая проблема, в защиту взаимодействия» (Поппер, Карл Р. Знание и психофизическая проблема: В защиту взаимодействия / Перевод с англ. — М.: Издательство ЛКИ, 2008. — 256 с.). Итак, что такое мыслительный процесс? Из определения мыслительного процесса становится понятно, что можно имитировать, а что нельзя. Мыслительный процесс состоит из следующих компонентов:

- 36 • Интенция, под которой понимаются вопросы, императивы, цели и установки;

поиск посылок и релевантных целей; рассуждение, и, наконец, рефлексия.

- Мысль есть любой элемент мыслительного процесса.

37 Отсюда понятно, что имитация мыслительного процесса – это совершенно нетривиальное дело, и некоторые аспекты этого мыслительного процесса мы имитировать не можем, в частности интенцию. И теперь мы можем определить, что такое естественный интеллект. Это система знаний, это набор познавательных способностей и высшие психические функции. К высшим психическим функциям относится рефлексия, интенция, инициатива и воображение. Естественно, что только некоторые из мыслительных способностей мы можем имитировать, в целом мы имитировать познавательный процесс не можем. К тому же он социален и взаимодействие со средой тоже входит в проблематику ИИ.

38 Я вам пытался сказать, но, наверное, не достаточно убедительно, следующее: познавательный процесс сводится не только к деятельности мозга. Это социальный процесс. Это одновременно запуск каких-либо инструментов, которые находятся у него в мозгу, это общение с коллегами. Поэтому я думаю, что имитация в точности мозговой деятельности – неразрешимая проблема. Если бы мы поняли, как мыслит человек, мы могли бы, наверное, стать Господом Богом. Поэтому эта проблема нерешаема, на мой взгляд.

39 **Алексеев А.Ю.:** Уважаемые коллеги. А вот до меня наконец-то дошло. Больше пятнадцати лет думаю о проблеме комплексного теста Тьюринга и вот сейчас дошло. Спасибо за это организаторам дискуссии, в особенности, модератору — Альберту Ефимову. На самом деле, никакого комплексного теста Тьюринга нет. В рамках шлейермахеровской герменевтики это очень даже хорошо объясняется: автор не владеет смыслом произведения, этот смысл выявляется лишь в ходе его последующей интерпретации. И интерпретации предыдущей интерпретации. Есть просто один единственный, точнее, единый тест Тьюринга. Можно назвать его как *истинный, настоящий тест Тьюринга*. Он и объединяет все возможные версии и версии версий игры в имитацию. Вряд ли он когда-то будет построен и, тем более, пройден. Потому что в познание интеллекта включено само познание, а это бесконечный процесс, познание познания. Вот только открыли нюанс, что-то, что отличает естественное от искусственного, только смогли его формализовать и запрограммировать — и вновь выявляется что-то новое, ранее не известное. Допустим, создали некоторый «нейроморфный интеллект» о котором многие говорят. Но вдруг оказалось, что формальная нейронная сеть, на самом деле, имеет субнейронный уровень. Только сделали методы настоящего, глубоко обучения и вдруг открыли феномен глобальной активизации нейронов, т.е. супернейронный уровень.

40 Таким образом, *тьюринговая методология* была и остается основной методологией искусственного интеллекта. Концептуальной методологией ИИ, как в отношении фундированности тестом Тьюринга всех возможных технологических приложений ИИ, так и в плане того, что позволяет нам рассуждать об ИИ на уровне смысла, концепций, значимостей, для чего все это надо. В этом плане так называемая. «*антитьюринговая методология ИИ*» (имеется и такой термин, возник среди журналистов) — это некоторый слепой, хаотичный, блуждающий от пиара к пиару путь развития ИИ. Из одного тупика — в другой.

41 Предложенная Альбертом Ефимовым *посттьюринговая методология* развития ИИ, однако, чрезвычайно интересна. Но не так, как это прозвучало вначале дискуссии: тест Тьюринга пройден, точнее, игра в имитацию сыграна в пользу машины. Этот формат методологии ИИ следует совместить с постнеклассической канонической рациональностью, предложенного Вячеславом Семеновичем Степиным. Это идеал, когда субъект познания и объект познания становятся неразличимыми в ходе самоорганизации системы, когда ментальные и телесные характеристики сливаются в этом процессе, выражая воплощенное

сознание и осознанную телесность, когда Я, то есть всё целостное то, что осталось «за скобками» феноменологического редуцирования моей компьютерной имитации, стал воспринимать себя как естественную искусственность и искусственное естество. Такая посттюринговая методология — это уже не тест Тьюринга, а тест киборга. То есть посттюринговый тест Тьюринга — это тест киборга, управляющего собственным сознанием посредством собственных нейральных коррелятов сознания. Будущее искусственного интеллекта – за киборгами. Я за то, чтобы им скорее стать.

References:

1. Alekseev A. Yu. Rol' kompleksnogo testa T'yuringa v metodologii iskusstvennykh obschestv // Iskusstvennye obschestva. 2011, T. 6, Vypusk 1-4 URL: <https://artsoc.jes.su/s207751800000051-5-1/>
2. Budushee iskusstvennogo intellekta: [Sbornik / AN SSSR; Red.-sost. K. E. Levitin, D. A. Pospelov]. // M. : Nauka, 1991. - 301
3. Finn V.K. Iskusstvennyj intellekt: Metodologiya, primeneniya, filosofiya. // Izd. stereotip. URSS. 2018. 448 s. ISBN 978-5-396-00861-8
4. Weizenbaum, Joseph Computer Power and Human Reason: From Judgment to Calculation. // New York: W.H. Freeman and Company. pp. 2, 3, 6, 182, 189.

Будущее искусственного интеллекта: тьюринговая или посттьюринговая методология?

Алексеев А. Ю.

*МГУ им. М.В. Ломоносова
Российская Федерация, Москва*

Ефимов А. Р.

*Сбербанк
Российская Федерация, Москва*

Финн В. К.

*Институт проблем информатики Федерального исследовательского центра
«Информатика и управление» РАН
Российская Федерация, Москва*

Аннотация

Идеи данной работы прозвучали на одноименной панельной дискуссии в рамках международного форума «Skolkovo Robotics & AI Forum», 16 апреля 2019 г., технопарк «Сколково». Обсуждалась роль, которую играет для методологии искусственного интеллекта тьюринговая игра в имитацию интеллектуального поведения (1950 г.) и тест Тьюринга (1960-е г. – наши дни). Контекстом дискуссий явились впечатляющие достижения современной робототехники, претендующей на всеобъемлющую имитацию человеческих форм коммуникации, не только вербальных, но и аудио-визуальных. Можно ли утверждать, что тест Тьюринга уже пройден и разработка интеллектуальных систем осуществляется в рамках некоторой посттьюринговой методологии, где снята острота проблемы различения естественного и искусственного интеллекта? Возможно и то, что подход к ИИ как к компьютерной имитации естественного интеллекта заводит нас в тупик и поэтому будущее ИИ следует связать с идеей усиления естественного интеллекта? Однако второй подход не снимает, напротив, обостряет проблему различения «естественное/искусственное»: что именно, как и зачем надо усиливать за счет компьютерного репродуцирования? Эти вопросы поставили участники дискуссии – д.ф.н. А.Ю. Алексеев (философский факультет МГУ), А.Р. Ефимов (руководитель лаборатории робототехники Сбербанка России), выступивший модератором дискуссии, д.т.н., проф. В.К. Финн (Институт проблем информатики ФИЦ «Информатика и управление» РАН), О.И. Пивоваров (генеральный директор «АйПи Лаборатория», организатор Открытой конференции по искусственному интеллекту OpenTalks.AI). Идеи последнего спикера в данной статье не воспроизводятся. Они были интересны, характеризовали особенности программирования российскими коллективами тьюринговой игры в имитацию, однако, вряд ли они значимы для философии, методологии и теории искусственного интеллекта.

Ключевые слова: искусственный интеллект, тест Тьюринга, частный тест Тьюринга, комплексный тест Тьюринга, интеллектуальные системы

Дата публикации: 19.12.2019

Ссылка для цитирования:

Алексеев А. Ю. , Ефимов А. Р. , Финн В. К. Будущее искусственного интеллекта: тьюринговая или посттьюринговая методология? // Искусственные общества. 2019. Т. 14. Выпуск 4 [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <https://artsoc.jes.su/s207751800007698-6-1/> (дата обращения: 28.09.2020). DOI: 10.18254/S207751800007698-6

User code: 0; Download date: 28.09.2020; URL - <http://artsoc.jes.su/s207751800007698-6-1/> All right reserved.