



Artificial societies. 2013-2020

ISSN 2077-5180

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

All right reserved

Issue 4 Volume 14. 2019

Technological background of Individuality of Human and its Computer imitation

A. Efimov

Sberbank

Russian Federation, Moscow

Abstract

In the article, the author analyzes the problems of human-computer communication in the context of artificial intelligence, augmented reality and a Turing methodology for comparing the capabilities of artificial and natural intelligence in a dialogue. It is argued that the tool with which the computer and humans communicate is of no less importance than the computer program with which the dialogue is conducted. As an example of the implementation of such visualization, the project “E.LENA” of a digital television anchor created at the Sberbank Robotics Laboratory was considered. The author gives his own version of Turing's specific test to study the features of non-verbal communication interfaces between a person and a computer. The key idea formulated by the author for further discussion is to provide artificial intelligence with visual avatar in order to avoid the alienation of man and computer and the loss of semantic information during its transmission from computer to person.

Keywords list (en): augmented reality, artificial intelligence, digital avatar, digital Turing double, Turing test, philosophy of artificial intelligence

Date of publication: 19.12.2019

Citation link:

Efimov A. Technological background of Individuality of Human and its Computer imitation // Artificial societies. 2019. V. 14. Issue 4 [Electronic resource]. Access for registered users. URL: <https://artsoc.jes.su/s207751800007645-8-1/> (circulation date: 28.09.2020). DOI: 10.18254/S207751800007645-8

Введение

Некоторые ошибочно полагают, что А. Тьюринг в своих работах, заложивших курс всей современной computer science, не упоминал «физических» роботов. Это не так, поскольку сам Тьюринг говорил о подобных машинах и считал, что делает это слишком преждевременно. В частности, в своем отчете 1948 года [8] он предполагает, что физическая интеллектуальная машина, созданная из моторов, телекамер и громкоговорителей, «блуждающая по полям» Англии, будет, «несомненно, создавать угрозу для окружающих» [8, стр. 3-23]. В силу этого несовершенства технологий в той области знаний, которую мы сейчас называем робототехникой, методология, предложенная им, была основана на человеческой речи, а точнее тексте. Другие естественные навыки человека было слишком сложно реализовать, тогда как обмен репликами посредством записок был намного доступнее для инженерного воплощения в его время. Тем не менее, прогресс компьютерных технологий приобрел такие формы, которые не мог предвидеть основатель искусственного интеллекта.

2 Изображения людей, животных или выдуманных существ на экране компьютера или кинозала не являются чем-то новым. Однако предыдущее поколение технологий компьютерной графики – от мультфильмов Диснея и до сложнейшей компьютерной графики современных голливудских студий – лишь подражало реальности, используя ее в качестве основы творчества. Художники, делающие компьютерную графику, создавали новую, виртуальную, реальность, которая настолько похожа на наблюдаемый нами мир, насколько похожи на него лучшие работы классических живописцев: они отражают не окружающий мир, но взгляд художника, влияющий на впечатления зрителей. Новое поколение технологий, которое появляется прямо сейчас, на наших глазах, делает доступным другой подход. Компьютерная графика становится не альтернативой реальному миру, но его живым дополнением: за основу изображения берется видео- или фотоизображение реальности, и уже на его основе создается видео или аудио запись. Итоговый видеофрагмент мог никогда не существовать в реальности, но выглядит он абсолютно реалистично. В этой связи совершенно правомерно вспомнить древнегреческую историю о двух художниках, Зевксисе и Парразии [3].

3 Зевксис с Парразием поспорили, кто лучше распишет стену храма. Собрался народ, вышли двое соперников, у каждого из них росписи были под покрывалом. Зевксис отдернул покрывало — на стене была виноградная гроздь, так похожая на настоящую, что птицы слетелись ее клевать. Народ рукоплескал. «Теперь ты отдерни покрывало!» — сказал Зевксис Парразию. «Не могу, — ответил Парразий, — оно-то у меня и нарисовано». Зевксис склонил голову. «Ты победил! — сказал он. — Я обманул глаз птиц, а ты обманул глаз живописца» [3].

4 На наш взгляд, виртуальную реальность, воссоздаваемую с максимальным правдоподобием компьютерной графикой (которая является лишь невероятно интеллектуальной кистью художника – специалиста по компьютерной графике или программиста), можно соотнести с мастерством Зевксиса, а технологии «дополнения/восполнения» реальности, связанные с обработкой изображения реальных объектов искусственными нейронными сетями, – с мастерством Парразия. Древние истории и здесь могут стать отправной точкой в последующем философском рассуждении.

5 Мы не предлагаем проводить сравнение объективной реальности, компьютерной графики, виртуальной и дополненной реальности. Скорее речь идет о сравнении способов коммуникации машины с человеком – насколько человек может доверять машине послать сообщение другому человеку. Может ли послание быть передано машиной так, как если бы

оно было передано человеком? Известный афоризм М. Маклюэна о том, что «средство коммуникации и есть сообщение» [5] мы можем дополнить и заявить, что «машина и есть средство коммуникации». М. Маклюэн не мог вообразить себе телевидение, в котором не было бы людей. Но сейчас это не только мыслимо, но уже реально.

6

Э.ЛЕНА – цифровой аватар, теледиктор

В начале 2019 года Сбербанк представил первого цифрового русскоязычного теледиктора Э.ЛЕНУ (Электронную *ЛЕНУ*), основанного на генерации реалистичного образа телеведущей. Использование технологий на основе искусственных нейронных сетей для улучшения изображений не является чем-то новым: в каждом смартфоне уже есть несколько приложений, которые модифицируют фотографии пользователей в соответствии с их самыми заветными чаяниями (убирают морщины, размыывают фон, исправляют цвета) практически в онлайн-режиме. Мы сами начинаем воспринимать окружающую реальность так, как если бы ее пропустили через фильтры популярных приложений, при этом обычная ясность взгляда на объективную реальность получила даже специальное обозначение (тег) *#nofilter*. Однако дополнение/воссоздание реальности является достаточно новой технологией: представленные пользователю объекты в видеофрагменте кажутся реальными, документальными, так как реалистично отражают знакомые пользователю сцены или объекты. В то же время то, что происходит с этими объектами, никогда не происходило в реальности и не может произойти. За рубежом данный феномен получил броское название *deepfake* и его стали связывать с отрицательными явлениями социально-политической жизни.

7

Сложно сказать, кто впервые предложил идею полной оцифровки актера или телеведущего, для этого нужно отдельное исследование поп-культуры и научной фантастики. Однако в явном виде идея полной оцифровки профессионального актера была представлена в малозамеченном фантастическом фильме «Конгресс» 2012 года.

8

Впервые цифровой теледиктор-двойник был представлен китайской компанией Sogou, которая разработала платформенное решение по заказу государственного новостного агентства Xinhua в ноябре 2018 года. Немногим позднее Сбербанк впервые самостоятельно разработал и представил эту технологию на русском языке. Цифровой теледиктор, представленный Сбербанком, может полностью автоматически озвучивать любой текст. Это позволяет использовать данное решение в качестве теледиктора новостей на корпоративном телевидении Сбербанка. В настоящее время с помощью этой технологии уже произведены десятки новостных сюжетов. Зрителями стали сотрудники и клиенты Сбербанка, которые просмотрели новости с участием цифрового теледиктора сотни тысяч раз с помощью различных каналов коммуникации. Рассмотрим эту технологию подробнее.

9

«Что у Э.ЛЕНЫ под капором?»

Голос Э.ЛЕНЫ представляет собой нейросетевой синтез речи на основе искусственных нейронных сетей глубокого обучения. Чтобы создать этот голос, необходимо было провести обучение нейронных сетей на базе специально подготовленных фонограмм голоса диктора (профессиональной актрисы) и разработать программное обеспечение, которое позволяет преобразовывать произвольный текст в речь.

10

Мимика Э.ЛЕНЫ формируется с помощью ансамбля искусственных нейронных сетей, предварительно обученных на специально подготовленных данных – видеоматериалах и 3D-сканах актрисы-прообраза (в настоящий момент голос и видео Э.ЛЕНЫ принадлежат

разным актрисам). В результате этих двухэтапных преобразований, происходящих без участия человека, получается мимика и речь цифрового теледиктора. Затем с помощью автоматизированного технологического инструментария и компонентов компьютерного зрения и систем распознавания речи производится обработка, в результате которой выявляются и устраняются ошибки, после чего реалистичное видео готово к использованию. Весь комплекс представляет собой целостное решение, основанное на нескольких независимых технологиях с компонентами искусственного интеллекта.

11 В настоящий момент сервис по преобразованию текста в видеофрагмент находится лишь в опытной эксплуатации у службы корпоративного телевидения ПАО «Сбербанк». Произведено порядка 50 различных новостных сюжетов с использованием данной программы. Текущая реализация Э.ЛЕНЫ имеет достаточное количество несоответствий: плохая синхронизация движения губ и произносимого текста, ограниченность поз, неестественный голос и т.п. Однако технологии очень быстро развиваются, и в самое ближайшее время многие компании и научные центры смогут продемонстрировать разработки высочайшего уровня.

12 Даже текущая реализация Э.ЛЕНЫ способна ввести ряд людей в заблуждение. Сбербанк провел опрос в сообществе, в которое входит 1.5 млн пользователей социальной сети «Одноклассники», о «происхождении Елены», представив посетителям своей группы два ролика одновременно: на одном был человек-теледиктор, на другом – цифровой теледиктор Э.ЛЕНА. Поразительно, что даже в текущей реализации более чем 25% опрошенных из 22 тысяч человек ошиблись (или проявили сомнение) при определении происхождения (цифрового или естественного) теледиктора.

13

Вербальная и невербальная коммуникация

Взаимодействие людей друг с другом основано на нашем взаимном понимании смысла коммуникации, отражающей не только интенциональность направленной на другого коммуникации, проявляемой в нашей речи и языке, но также контекст взаимодействия, который может быть географическим, временным или смысловым. При взаимодействии друг с другом мы также учитываем невероятное количество социокультурных особенностей друг друга (к примеру, в академической среде используем иные выражения, нежели совершая покупки на рынке). В общем виде, чтобы машина (компьютер, робот) понимала человека, необходимо обеспечить понимание всех трех аспектов смысла, который мы вкладываем в речь: язык, контекст и культура. Поэтому подход к изучению искусственного интеллекта, сфокусированный исключительно на обработке естественных языков, представляется недостаточным для понимания смысла, вложенного в коммуникацию.

14 Все три модальности понимания смысла (язык, культура, контекст) отражаются не только в буквальном значении слов, но и, по мнению Д. Эверетта [7], с помощью имплицитного культурного знания, которое он называет «темной материей». Эта темная материя может состоять не только из смысловых конструкций, образованных словами и выражениями, но и, например, из жестов, сопровождающих локутивные действия.

15 Милен Долар в своей книге приводит пример того, как Станиславский давал задание своим ученикам подготовить пятьдесят различных способов произнести фразу «завтра вечером», вкладывая в нее различные намерения [4]. Говорящие включают мимику лица, жесты и интонацию в синтаксическую структуру предложения и используют их как пояснения, указывающие на наличие имплицитной информации, содержащейся в культуре или личном опыте говорящего и слушающего. Д. Эверетт справедливо замечает, что «язык никогда не выражает всего, культура заполняет эти пропуски» [7, стр. 272]. Традиционный подход к исследованию искусственного интеллекта, основанный на тексте, а по сути, на «телетайпных»

сообщениях, унаследованных из эпохи аналоговой электроники, игнорирует «темную материю» коммуникации, поскольку интерпретация сообщения (по Эверетту) направлена не только на вербальное рассуждение, но и на жесты и мимику.

16

Цифровой теледиктор как инструмент для изучения коммуникации человека и машины

С того момента, как А. Тьюринг предложил заменить вопрос «может ли машина мыслить?» на игру в имитацию, которую он предлагал проводить с помощью обмена «записочками» или телетайпными сообщениями, исследователи искусственного интеллекта, по сути, не обращали никакого внимания на то, каким образом сообщения передаются между судьей и испытуемыми через так называемую «стену Тьюринга», отделяющую участников игры в имитацию. Сотни научных и популярных работ, посвященных искусственному интеллекту, обходили стороной вопрос «невербальной» коммуникации с машиной.

17 Одним из немногих исключений является советский научно-популярный фильм «Кто за стеной?» 1977 года, в котором «стена Тьюринга» стала видеостеной. По мнению авторов настоящей статьи, Э.ЛЕНА могла бы стать новым инструментом для изучения проблематики человеко-машинного взаимодействия и искусственного интеллекта за счет расширения методики А. Тьюринга, которая позже получила название «тест Тьюринга». Современный прогресс в области создания «дополненной реальности», появление цифровых аватаров, теледикторов типа Э.ЛЕНА ставит перед нами другой важный вопрос: может ли машина создавать такую же интерпретационную основу речи, как и человек, используя не только определенный набор слов для выражения мысли или намерения, но и невербальные коммуникации – мимику и жесты? Будет ли обладать машина таким арсеналом для локутивного акта, каким обладает человек? Или же, подняв стену Тьюринга, мы увидим лишь микросхемы и аккумуляторы, которые нам безошибочно укажут на обман.

18 Технологии создания *дополненной цифровой реальности* дают нам возможность сформировать новый *несовершенный частный тест Тьюринга* в терминологии А.Ю. Алексеева [1]. Частным несовершенным тестом Тьюринга называется испытание, которое сфокусировано на проверке лишь одного компонента исходного теста Тьюринга. В данном случае предложенный частный тест Тьюринга (ЧТТ) направлен на проверку невербальных коммуникационных возможностей компьютерных программ. По мнению А. Алексева, ЧТТ описывается следующими компонентами: предметом тестирования, реализационной схемой, тестовыми вопросами-ответами. Кроме того, А. Алексеев предлагает дополнять описание самого тестирования дискуссией (подобно тому, как А. Тьюринг подошел к разбору возражений относительно оригинального теста) и описанием социокультурных последствий. При описании предлагаемого частного теста Тьюринга «Э.ЛЕНА» мы воспользуемся предложенным подходом.

19 По сути, Э.ЛЕНА представляет собой симуляцию виртуального мира телевизионной студии. Задача телевизионной студии как средства массовой информации – сформировать у своей аудитории определенную картину мира. В предлагаемом ЧТТ «Э.ЛЕНА» предметом тестирования является способность человека воспринимать информацию, предлагаемую цифровым телеведущим, ведущим вещание из цифровой телестудии. В самом деле, Э.ЛЕНА (а точнее программный комплекс, создающий ее) преобразует текстовую информацию в аудиовизуальный формат, который, по замыслу создателей системы, должен быть аналогичен формату телевизионной новостной студии. В предлагаемом тесте мы определяем, как люди воспринимают информацию от цифровых теледикторов, и есть ли какая-либо разница в восприятии человеком информации, переданной ему теледиктором-человеком или цифровым

теледиктором.

20 Как и исходный тест Тьюринга, тест Э.ЛЕНЫ имеет бинарный результат. Если наблюдатель воспринимает информацию не хуже, чем при просмотре новостей, озвученных реальной ведущей, то тест пройден. Если же наблюдатель при просмотре новостей, озвученных роботом, воспринимает информацию хуже, то тест не пройден. Конкретную имплементацию теста Э.ЛЕНА назовем для краткости «эксперимент».

21

Реализационная система

Уточним терминологию, используемую во всех вариантах дизайна эксперимента. Объектом исследования является *наблюдатель*, в терминологии А. Тьюринга и А.Ю. Алексеева – судья (С). Э.ЛЕНА является инструментом («средством коммуникации» в терминологии М. Маклюэна), который используется для изучения реакций С на получаемую информацию («послание» в терминах М. Маклюэну).

22 Э.ЛЕНА – цифровой аватар, теледиктор – компьютерная программа, преобразующая составленный экспериментатором текст в видеоряд. Преобразование происходит мгновенно (1 секунда видео генерируется менее чем за одну секунду).

23 Судья (С) – человек, последовательно просматривающий видеофрагменты, среди которых есть и Э.ЛЕНА, выполняющая роль теледиктора. Предметом тестирования (исследования) являются реакции С на Э.ЛЕНУ: ошибка восприятия информации и отторжение (принятие) Э.ЛЕНЫ в качестве теледиктора (источника информации). С – главный объект эксперимента. Человек (Ч) – лицо, выполняющее роль *теледиктора* в различных вариантах эксперимента. Теледиктор (Т) – роль, которую может выполнять Ч или Э.ЛЕНА.

24 В ходе конкретного эксперимента С предлагается посмотреть видеоролики, на которых люди рассказывают о себе или отвечают на вопросы анкеты. На роликах Т, включая Ч и Э.ЛЕНУ в похожей одежде и в одинаковой студии, озвучивают текст, который у них примерно совпадает по тематике и стилистике. После чего С предлагается дать оценку качеств этих Т. Если С никак не обособляет Э.ЛЕНУ от остальной группы участников, считаем *тест пройденным*.

25

Тестовая программа

ЧТТ «Э.ЛЕНА» предлагается в двух вариантах экспериментов. Вариант 1. «Найди робота». В этом варианте эксперимента С последовательно просматривает несколько видеофрагментов, в которых роль Т могут играть Ч или Э.ЛЕНА. Каждый видеофрагмент длится не более 25 секунд (в терминологии новостной журналистики такой фрагмент называется «синхрон») (Школа журналистики, 2017). В соответствии с правилами формирования новостных сюжетов таких фрагментов не должно быть более четырех. Именно поэтому в предложенном тесте количество роликов различных теледикторов, среди которых Э.ЛЕНА, ограничено именно этим числом. Фон для всех видеофрагментов одинаков. Все тексты Т имеют одинаковую тематическую направленность (культура, спорт, погода и т.п.). Задача С – определить, в каком из фрагментов роль теледиктора выполнила Э.ЛЕНА. Тест считается пройденным, если вероятность определения Э.ЛЕНЫ во всех видеофрагментах не составляет более 50%.

26

Вариант 2. «Счастье – это когда тебя понимают». В этом варианте эксперимента С

просматривает несколько видеофрагментов, в которых роль теледикторов могут играть Ч или Э.ЛЕНА. Каждый ролик длится не более 25 секунд. Всего демонстрируется четыре видеофрагмента различных Т (среди которых несколько Ч и Э.ЛЕНА). Фон для всех видеофрагментов одинаков. Все тексты теледикторов имеют одинаковую тематическую направленность (спорт, культура, погода). Задача С – определить для каждого из просмотренных видеофрагментов ряд ключевых фактов или мнений, которые ясно выражены в данном отрывке. Опрос С проводится сразу же после просмотра. Тест считается пройденным, если средняя ошибка С при ответах на вопросы после просмотра видеофрагментов с Ч такая же, как при ответах на вопросы после просмотров роликов Э.ЛЕНЫ.

27 Каждый из вариантов эксперимента проводится сериями, в которых роли С выполняют разные Ч. Следует провести не менее 10 серий с каждым С.

28

Дискуссия и возможные возражения

Ключевое возражение, которое можно сделать против разработки предложенного ЧТТ, состоит в следующем: *данная разработка не относится к области искусственного интеллекта, а скорее представляет собой новый вид компьютерной графики*. В данном возражении смешивается технология, которую мы используем для получения цифрового аватара (в частности, теледиктора), и ее значимость для реализации ТТ в общем смысле. Игроки в имитацию могли бы использовать образ цифрового теледиктора для того, чтобы успешнее вводить в заблуждение С при проведении классического ТТ. С другой стороны, следуя Эверетту, мы указываем, что для понимания смысла невербальные коммуникации могут иметь значение не меньше чем сама речь – именно этим объясняется широкое распространение видеосвязи для общения среди молодого поколения. Как авторы указывали выше, А. Тьюринг предлагал использовать телетайп только потому, что в его время невозможно было даже представить что-то подобное Э.ЛЕНЕ. Сейчас это стало уже реальностью, которую невозможно игнорировать в исследованиях взаимодействия человека и машины.

29 Другое возражение может быть основано на финальном, успешном прохождении описанного теста «Э.ЛЕНА». Можно сказать, что в этом случае ИИ, ведущий диалог с С «из-за стены», будет всего лишь отражением. Правда, не С, но того образа, который послужил основой для Э.ЛЕНЫ. Но мы никогда не утверждали, что данный тест является «общим» тестом на интеллект. Он всего лишь «уравнивает» невербальные возможности человека и машины. Если до появления Э.ЛЕНЫ и подобных ей систем у машин не было возможностей воздействия на интерпретационные каналы человека в полном объеме, то теперь они могут быть исследованы и последовательно реализованы.

30

Заключение.

Безусловно, мы еще очень далеки от фантастической идеи А. и Б. Стругацких о том, что «модели на базе собственных дублей умел создавать любой сотрудник, защитивший магистерскую диссертацию» [6, стр. 74]. Сейчас процесс создания полного цифрового двойника теледиктора занимает несколько месяцев работы серьезного междисциплинарного коллектива инженеров и специалистов по видеообработке. Но прогресс идет чрезвычайно быстро. Опыт эксплуатации Э.ЛЕНЫ в Сбербанке показывает, что профессия теледиктора может уйти в прошлое в течение нескольких лет. Однако мы не знаем, поможет ли цифровой аватар людям лучше понимать искусственный интеллект. Мы лишь знаем точно, что это поможет нам меньше бояться неизбежного будущего.

References:

1. Alekseev A.Yu. Kompleksnyj test T'yuringa: filosofsko-metodologicheskie i sotsio-kul'turnye aspekty. // Moskva: IInteLL, 2013
2. Gavrilov K. Kak delat' syuzhet novostej? i stat' mediatvortsom. // SPb: Amfora, 2007
3. Davydova L.I., Kon'kova G.I., Chubova A.P. Antichnye mastera. Skul'ptory i zhivopistsy. // L: Iskusstvo, 1986.
4. Dolar M. Golos i nichego bol'she. // M.: Izd-vo Ivana Limbakha, 2018
5. Maklyuehn M. Ponimanie Media: vneshnie rasshireniya cheloveka. // M.: Giperboreya; Kuchkovo pole, 2007.
6. Strugatskij A.N., Strugatskij B.N. Ponedel'nik nachinaetsya v subbotu: Skazka dlya nauchnykh rabotnikov mladshego vozrasta. // M.: Detskaya literatura, 1965.
7. Ehverett D. Kak nachinalsya yazyk. Istoriya velichajshego izobreteniya. // M.: Al'pina Non Fikshn, 2019
8. Turing Alan Intelligent Machinery // 1948. In Cybernetics: Key Papers, eds. C. R. Evans and A. D. J. Robertson, University Park Press, Baltimore Md.and Manchester (1968).

Технологические предпосылки неразличимости человека и его компьютерной имитации

Ефимов А. Р.

Сбербанк

Российская Федерация, Москва

Аннотация

В статье автор анализирует проблематику коммуникации человека и компьютера в контексте разработок в области искусственного интеллекта, дополненной реальности и тьюринговой методологии сравнения возможностей искусственного и естественного интеллекта в диалоге. Утверждается, что инструмент, с помощью которого ведется коммуникация компьютера и человека, имеет не меньшую важность, чем компьютерная программа, ведущая диалог. В качестве примера реализации такой коммуникации рассмотрен проект «Э.ЛЕНА» – цифровой теледиктор, созданный в Лаборатории робототехники Сбербанка. Автор предлагает собственный вариант частного теста Тьюринга для изучения особенностей невербальных коммуникационных интерфейсов между человеком и компьютером. Ключевая мысль, сформулированная автором для дальнейшей дискуссии, состоит в наделении искусственного интеллекта визуальным образом для того, чтобы избежать отчуждения человека и компьютера и потери смысловой информации при передаче от компьютера к человеку.

Ключевые слова: дополненная реальность, искусственный интеллект, цифровой аватар, цифровой двойник Тьюринга, тест Тьюринга, философия искусственного интеллекта

Дата публикации: 19.12.2019

Ссылка для цитирования:

Ефимов А. Р. Технологические предпосылки неразличимости человека и его компьютерной имитации // Искусственные общества. 2019. Т. 14. Выпуск 4 [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <https://artsoc.jes.su/s207751800007645-8-1/> (дата обращения: 28.09.2020). DOI: 10.18254/S207751800007645-8