

# ОБЩИЕ ВОПРОСЫ ОБЩЕГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Алексеев Андрей Юрьевич

**Аннотация.** Проект общего искусственного интеллекта, который обеспечивает программно-аппаратную реализацию обширного спектра когнитивных феноменов, удобно изучать на основе исследований комплексного теста Тьюринга. Этот тест ориентирован на решение главного вопроса философии искусственного интеллекта «Может ли компьютер всё?». Квантификатор всеобщности в данном суждении пробегает по частным тьюринговым тестам, которые предназначены для поиска ответов на вопросы: может ли компьютер мыслить, понимать, жить, творить, осознавать, любить, дружить и пр.? Интегральное решение этих тьюринговых вопросов, приводящие к пониманию системы ИИ как системы общего ИИ, целесообразно искать путем экспликации базовых функций комплексного теста Тьюринга: коммуникативной, интеррогативной, организационной, дефинитной, критической, конструирующей и конститутивной функций.

**Ключевые слова:** общий искусственный интеллект, комплексный тест Тьюринга, формальное определение общего ИИ

**Для цитирования:** Алексеев А.Ю. (2023). Общие вопросы общего искусственного интеллекта. – *Исследования в цифровой экономике*. №3, С. 6–37. [DOI: 10.24833/14511791-2023-3-6-37](https://doi.org/10.24833/14511791-2023-3-6-37)

## **Об авторе:**

**Алексеев Андрей Юрьевич,**

ГАУГН, философский факультет, 119049, г. Москва, Мароковский переулок, дом 26.  
РУДН, Инженерная академия, департамент механики и процессов управления, 115419, г. Москва, ул. Орджоникидзе, д. 3  
[aa65@list.ru](mailto:aa65@list.ru)

## Введение.

### Основные идеи проекта общего искусственного интеллекта

Словосочетание «общий искусственный интеллект» («artificial general intelligence», AGI) вряд ли можно использовать в роли термина в национальных и ведомственных стандартах для осуществления соответствующих нормативных актов. Это понятие в последние годы часто применяется в проектах, получивших собственное имя «Чат GPT» и связывается с продуктами компаний OpenAI, DeepMind, Anthropic и подобных компьютерных фирм. Тем не менее AGI до сих пор принадлежит сфере научной фантастики для обозначения технических систем, способных выполнять любую когнитивную задачу животного, человека, общества. Постулируется то, что системы *общего ИИ* способны реализовывать когнитивные функции, которые относятся не только к сфере *другого сознания*, но и к *иному* и *чужому сознанию*. То есть эти функции сложно вообразить. Однако, как предполагается AGI-пропонентами, их можно помыслить хотя и в затейливом математическом формате. Именно в таком математическом смысле осуществляются попытки рационального понимания *общего ИИ* в рамках замечательных ежегодных международных конференций «AGI» (см. сайт конференции [13] и сборник первой конференции 2008 г. [11]).

В настоящей работе мы не будем заниматься поиском математической «Чаши Грааля» *общего ИИ*. Этим озабочены участники конференций AGI: найти такую математическую формулу и реализовать соответствующую компьютерную программу, которые способны работать в самых разнообразных областях когнитивной активности. Наша задача менее амбициозна: исходя из общей позиции, опираясь на философско-методологические, холистические, концептуальные, смысловые формы анализа изучить самые общие вопросы *общего ИИ*.

Первым из ряда подобных вопросов является определение понятия. Под *общим искусственным интеллектом* мы будем понимать философско-методологические, научно-теоретические и инженерно-технологические исследования, ориентированные на построение и применение компьютерных репликаций (имитаций), репрезентаций (моделей) и репродукций (воспроизведений) когнитивных феноменов широкого спектра жизненных, психических, личностных и общественных проявлений.

Общие вопросы о путях развития этого проекта решаются на уровне философии и методологии искусственного интеллекта [5]. Ранее в [14] предлагалось использовать т.н. «общефункционалистский подход» для концептуальной организации проекта *общего ИИ*. Общий функционализм продолжает две главные идеи машинного функционализма Х. Патнэма. Первая идея связана с функциональным гомеоморфизмом естественных и искусственных систем: причинные связи мозга и психики подобны отношениям вычислительного устройства к логике его функционирования. Вторая идея – множественной реализации – утверждает об инвариантности когнитивной функции к субстрату реализации:

реализация той же функции может быть осуществлена биологическим мозгом человека либо компьютерной системой искусственного интеллекта либо тинной марсианина (как мозгом агента, обладающего *иным и чужим сознанием*).

Продолжая классический машинный функционализм мы, однако, в нашей общей версии функционализма не вгоняем психику человека в детерминанты программируемой логики работы универсальной вычислительной машины человека. Такую явно антигуманистическую позицию выражал Х. Патнэм в ранний, оптимистичный период своего творчества. Позднее, однако, он осуществил радикальный скептический поворот: когнитивная физика столь же невозможна, как социальная физика О. Конта [14]. На наш взгляд, такой поворот надо смягчить. В самом деле, основным в машинном функционализме является использование идеи индетерминированной машины Тьюринга (1936 г.) как технического средства реализации когнитивной функции. Но Х. Патнэм неправильно ей воспользовался (в 1960 г.): он не учел того, что тест Тьюринга (1950 г.) не отождествляет программу со способом ее реализации, а всего лишь *оценивает* с позиции наблюдателя (судьи, другого человека) *возможность* инвариантности функций, реализуя их посредством программы квазиалгоритмического формата. Общий функционализм ИИ – это тестовый функционализм, основанный на идее тестирования, а не на идее отождествления.

Общий функционализм ИИ «стоит на трех китах»: на собирательном, определительном и наблюдательном функционализмах. *Собирательный функционализм* обеспечивает сбор, идентификацию, координацию, формализацию, систематизацию, унификацию и кодификацию всевозможных функционалистских теорий. На самом общем уровне он выполняет задачи философского анализа (точнее, аналитической философии): распутывание «клубков» противоречий как конфликта разнообразия теоретических интерпретаций. В конкретных условиях «собираения» решается стандартная библиографическая задача систематизации функционалистских теорий, эмпирических исследований и их методологических концептуализаций.

*Определительный функционализм* выявляет т.н. «главные» функционалистские характеристики, отношения, закономерности, причинные связи. Главными их называют потому, что составляющие определительных функций отвечают за *сущностную, каузально обусловленную* реализацию, которая обеспечивает достижение целей при условии инвариантности когнитивных феноменов относительно структуры и субстрата их реализатора. Например, главным является отношение множественной реализации феномена «нормальное физиологическое функционирование человеческого сердца», при котором сердце нормально работает даже в условиях соответствующей патологии, потому что за перекачку крови отвечает искусственный клапан.

*Наблюдательный функционализм* оценивает с позиции абстрактного или конкретного наблюдателя (исследователя, interrogator'a) различные статусы технологических реализаций и, в ходе концептуального обобщения, выделяет

категориальные параметры проекта общего ИИ. Обобщение предполагает категоризацию характеристик общего функционализма на следующих уровнях, традиционных для философского анализа: онтологическом, эпистемологическом, логическом, лингвистическом, аксиологическом, эстетическом, этическом и праксеологическом. Именно поэтому, за счет философских обобщений, данный проект именуется «общим ИИ», а не из-за того, что математические манипуляции позволяют применять одинаковые программные модули с точностью до означивания параметров, скажем, и в задачах распознавания образов и в задачах формирования траектории движения шагающего робота.

### Состав проекта общего ИИ

Проект общего ИИ в условиях такой философской категоризации целесообразно разделить на пять составных частей – подпроектов. Каждый современный подпроект исследований ИИ, как правило, имеет транснациональную значимость и фундируется особым классом когнитивных феноменов (функций), компьютерную реализацию которых эти проекты осуществляют. Это следующие проекты: искусственной жизни, искусственного сознания (мозга), искусственной личности, искусственного общества.

Проект *искусственной жизни* ориентирован на компьютерную реализацию витальных феноменов и отвечает на вопросы: может ли компьютер «жить»? По крайней мере, о правдоподобности имитации программно-аппаратной системой феномена зарождения, развития, функционирования и умирания искусственной системой с учетом генетических механизмов.

Задачи компьютерной реализации ментальных (психических) феноменов формируют проект, который звучит двояко, хотя значит одно и то же: может ли компьютер сознать? Натуралистические тренды этот проект называют как «*искусственный мозг*», метафизические ориентиры этот проект обозначают словосочетанием «*искусственное сознание*».

Проект «*искусственная личность*», имитирующий персонологические феномены (личность, индивидуальность, свобода, Я, смысл, творчество) был первым проектом в сфере ИИ. Дело в том, что основная работа для философии искусственного интеллекта – статья А. М. Тьюринга 1950 г. «Вычислительные машины и интеллект» – демонстрировала возможности универсальной цифровой вычислительной машины и обучаемой нейронной сети (машины-ребёнка) играть в игру имитации различных аспектов межличностной коммуникации.

Проект «*искусственное общество*» программирует социальные феномены: может ли компьютер быть обществом? Это применение компьютерного инструментария необходимо для моделирования социального поведения, точнее, для изучения и воспроизведения феноменов intersубъективной реальности – дружбы, любви, ненависти и др.

Наконец, совсем недавно, с октября 2019 г., когда корпорация Facebook была переименована в корпорацию Meta, открыто заявил о себе проект «искусственный мир». «Может ли компьютер всё?» – это главный вопрос нового проекта. Концепция вычислимости участвует во всех самых широких, высоких и глубоких, во всех самых низших и мелких проявлениях компьютерной имитации мира. Основной вопрос Метаверса и основной вопрос комплексного теста Тьюринга совпадают.

### Базис обобщения подпроектов общего ИИ

Отметим, однако, что обозначенные выше подпроекты, несмотря на разнообразие когнитивных феноменов, базируются на проекте искусственного интеллекта. Вначале необходимо интеллектуальное оформление того феномена, который обозначает «жизнь», «сознание», «личность» и пр. и лишь после такой предваряющей подпроект работы возможно отождествление функциональных репликаций, репрезентаций и репродукций когнитивного феномена собственно с самим феноменом. Например, базисом «искусственного сознания» является проект «искусственного интеллекта» в части программирования представлений о сознании. Никакого «искусственного сознания» как такового в этом подпроекте. Есть лишь терминологическая путаница.

Чтобы избежать номинативных затруднений в [2] в роли базового когнитивного феномена, который является общим интегратором всех базисных оснований подсистем функциональных отношений отмеченных выше подпроектов предлагалось выбрать не концепт «интеллекта» (совместно с его коррелятами «данных», «знаний» и «смыслов») и не «боль», понятие которой, как правило, используют для демонстрации собственных теоретических положений философы «*of mind'a*». Более продуктивным оказывается феномен потребности. Он обеспечивает широту охвата когнитивных функций в проекте общего ИИ, так как часто служит внешним оформлением внутренних процессов мышления, понимания, осознания, любви и пр. Самое главное, в понятии потребности фиксируется причинная зависимость между «должным» и «сущим», то, что *надо* и то, что *есть*. На основании функционалистского моделирования потребности возникает так называемая «искусственная потребность». Именно данное понятие и является, как доказывается в [2], претендентом на роль главной дефиниции и концептуальной основы интеграции подпроектов общего ИИ.

В настоящей работе мы не будем акцентировать внимание на методологии главного когнитивного феномена проекта общего ИИ. Возможно, что этим феноменом следовало бы выбрать не «потребность», а «творчество» (как было предложено в [7]) и, после данного выбора, далее целесообразно бы было прорабатывать подпроекты в стиле бергсоновской эволюционной эпистемологии. Вполне возможно, что в проекте общего ИИ необходимо учитывать класс ког-

нитивных феноменов. Каждый из феноменов способен играть роль интегратора подпроектов общего ИИ и зависит от конкретных предпочтений судьи (наблюдателя) комплексного теста Тьюринга.

### Формальное определение общего искусственного интеллекта

Комплексный тест Тьюринга целесообразно полагать настоящим, истинным, правильным тестом Тьюринга. На самом деле часто употребляемое в сегодняшних научных и околonaучных дискуссиях понятие «тест Тьюринга» не является чем-то конкретным, эмпирически воплощаемым. Тест Тьюринга (ТТ) – это собирательное понятие для исследований системы искусственного интеллекта на предмет возможности программно-аппаратной реализации самых различных когнитивных функций. Комплексный ТТ состоит из сотен версий оригинальной тьюринговой «игры в имитацию интеллекта» (1950 г.) и многочисленных версий этих версий. Сегодня, спустя три четверти века после рождения идеи тьюринговой игры в имитацию интеллекта, эти частные тесты Тьюринга раскрывают достаточно сложную когнитивную феноменологию компьютерного мира современной культуры. Может ли компьютер мыслить, понимать, творить, сознавать, самосознавать, любить, быть личностью и пр. Короче, может ли компьютер *всё*? Квантификатор всеобщности пробегает по всему множеству тьюринговых тестов, оценивая возможности и перспективы программно-аппаратной имитации этого весьма обширного и многомерного спектра когнитивных феноменов. Именно такой *всеобщий* охват позволяет утверждать о том, что компьютерная система, реализующая комплексный ТТ, воплощает концепцию *общего* искусственного интеллекта.

Идея комплексного теста Тьюринга прорабатывалась автором на протяжении двух десятилетий. Она подробно изложена, например, в [5, 1]. В одной из последних работ [6] доказывалось, что комплексный тест Тьюринга является формальным определением ИИ, аналогично тому, что машина Тьюринга является формальным определением понятия «алгоритм». В самом деле, понятие комплексного теста Тьюринга способствует компактному и ёмкому формальному определению ИИ: *искусственный интеллект – это квазиалгоритмическая реализация комплексного теста Тьюринга*. Пожалуй, значимость идеи комплексного теста в наибольшей степени подчеркивается дистинкцией между логически достаточными и логически необходимыми условиями развития ИИ. Логически достаточные условия ИИ – это те, которые способствуют развитию ИИ. К ним относится трудно регистрируемая совокупность направлений, мероприятий, проектов и др. деятельности, широко представленная в [4]. Логически достаточных условий развития ИИ много. К ним относится по сути все те исследования в методологии ИИ, которые по крайней мере не вредят развитию ИИ, а способствуют теоретическому знанию, инженерным проектам и попыткам связать воедино теорию с практикой в сфере ИИ.



Логически необходимых условий развития намного меньше. Это те условия, при отсутствии которых невозможно само развитие ИИ. Тест Тьюринга и есть такое необходимое условие. Дело в том, что работы по созданию и развитию систем ИИ должны предваряться убеждением разработчика в том, что компьютер (машина) может мыслить, либо когда-то сможет мыслить, либо никогда не сможет мыслить. Слово «мыслить», как мы отмечали выше, репрезентирует и фундирует много других когнитивных феноменов – понимание, сознание, творчество, любовь... И то, что касается этих феноменов, входит в компетенцию теста Тьюринга. Для исключения путаницы с понятиями «ИИ» / «общий ИИ» предлагается уточнять: не «тест Тьюринга», но «комплексный тест Тьюринга». С учетом этого уточнения получается то, что *комплексный тест задает необходимые условия развития общего ИИ*, так как поднимает принципиальные вопросы о компьютерной реализации разнообразного спектра когнитивных феноменов.

При обнаружении невозможности реализовать тот или иной частный тест Тьюринга, то есть при отсутствии положительного решения по поводу концептуальной реализации выбранной версии частного теста Тьюринга, принципиально нецелесообразно приступать к разработке конкретного подпроекта ИИ, воплощающего данный частный тест Тьюринга. И это чрезвычайно важно для цифровой экономики в целом. Если, например, мы докажем невалидность понятия «когнитом», то это значительно сократит расходование ресурсов на тупиковую ветвь развития ИИ в части подпроекта «искусственный мозг».

Исходя из важности изучения логически необходимых условий ИИ, существующей сложности организации комплексного ТТ (сотня тьюринговых тестов!) и гиперсложности исследования вопросов метауровневого порядка относительно уровня исследования достаточных условий ИИ, не подлежит сомнению актуальность и важность изучения общих вопросов по поводу разработки и реализации проекта общего ИИ на базе концептуальных оснований, которые предоставляет комплексный тест Тьюринга.

### **Общие функции комплексного теста Тьюринга в проекте общего ИИ**

В литературе выделяется не менее сотни крупных версий теста Тьюринга. Ряд версий мы ранее абстрагировали, обобщили в [5, 1] и предложили изучать двадцать одну версию как наиболее репрезентативную подборку т.н. совершенных частных тестов Тьюринга. Идеи совершенных тьюринговых тестов сюръективно покрываются концептами А. Тьюринга, изложенными в основополагающей для искусственного интеллекта работе [15]. Напомню то, что в оригинальном тесте Тьюринга, хотя и употребляется много слов по поводу имитации «mind», на самом деле раскрывается игра в имитацию гендера, совершаемую и человеком и компьютером как цифровым «двойником» человека.

Для целей нашей работы это не существенно, тем не менее замечание раскрывает сложность формулировки: чем же на самом деле занимается истинной ИИ, сформулированный в тесте Тьюринга (1950 г.) вне связи с различными акциденциями ИИ как слабый, сильный, узкий, глобальный, доверенный, и наконец, общий ИИ? Получается, что на заре ИИ предметная область «интеллекта» лежала в более общих областях, нежели чем «интеллект диалогового типа». Это еще раз подчеркивает вариативность базового понятия для общего ИИ: потребность, творчество, гендер. Какие иные возможны репрезентанты?

Изучение общих вопросов общего ИИ удобно организовать путем изучения функций тестового функционализма, общих для частных тестов Тьюринга [1]. К этим *функциям обобщения* (интеграции) частных тестов относятся коммуникативная, интеррогативная, дефинитная, критическая, конструкторская, организаторская и конститутивная функции. В настоящей статье эти функции интерпретируются в формате вопросов, которые необходимо решить для реализации комплексного теста Тьюринга и, тем самым, для раскрытия базовых идей тестового функционализма общего ИИ. Это и позволит изучить общие вопросы общего ИИ.

**1. Коммуникативный вопрос общего ИИ** обеспечивает предварительную междисциплинарную координацию специалистов, которые участвуют в создании проекта ИИ. Тест Тьюринга вводит метафоры, простые категории и отчетливые образы, которые понятны и физикам и лирикам, философам и программистам, лингвистам и юристам, экономистам и химикам и мн.др. Для организации междисциплинарного взаимодействия целесообразно применять т.н. трехмерный язык когнитивных терминов [3]. Имеются две интерпретации такого 3D-языка: наивная и зрелая. Наивная трактовка выделяет единство трех семантик когнитивного термина в соответствии с мотивами, навеянными т.н. «философией сознания»: психическую, физическую и вычислительную. Наивная трактовка проработана Д. Чалмерсом в решении психофизической проблемы с позиции т.н. «натуралистического дуализма», но с весьма существенным ограничением – он игнорировал вычислительный контекст, который, по нашему мнению, выделяет третью семантическую метрику. Более зрелая форма 2-дименсионализма Д.Чалмерса [12] определяет двумерную семантику исходя из кантовских понятий априори (первое измерение) и понятий апостериори (второе измерение). Первое измерение как правило охватывает феномены сознания, субъективную реальность. Второе измерение – натуралистические феномены объективной реальности. Но опять Д. Чалмерс даже в более строгой трактовке двумерности игнорировал погруженность исследователя в контекст электронной культуры, в вычислительный контекст. Снова у ведущего мирового метафизика отсутствует понятие вычислимости. Мы предлагаем ориентироваться на реалии искусственного интеллекта и на функционализм как нативную методологию ИИ [15]. То есть мы имеем одно измерение как *когницию*. Это априорные понятия, феномены сознания. Второе измерение рассматрива-



ется как *реализация*, т.е. апостериорные понятия, феномены, которые можно потрогать, пощупать, увидеть, услышать, сфотографировать – примерно таким образом выражался В.И. Ленин, характеризуя материалистическую трактовку реальности. Третье измерение – это собственно, то, что конституирует понятие «измерения». Феномены третьего измерения представляются как *компьютации* – вычислительные алгоритмы, а также весь комплекс программных, информационных, технических, лингвистических и других средств информационных технологий. Трехмерная семантика, выражающая смысловое триединство когнитивного, технического и алгоритмического представляется конструктивной основой языка коммуникации, так как метафоры, образы, категории общего ИИ нанизывает на достаточно отчетливый концептуальный каркас.

**2. Интеррогативный вопрос общего ИИ** характеризует тьюринговые вопросы, которые задает тьюринговый судья (interrogator, наблюдатель) х-системе на предмет ее у-способности с позиции собственного обладания судьей z-компетенцией. То есть в данном пп. изучается «вопрос о вопросах», точнее, «вопрос о тьюринговых вопросах». Важность правильного тьюрингового вопрошания к системе ИИ имеет давнюю историю, начиная со статьи А.Тьюринга 1950 г. Немного позже она проявилась в изучении вопросов к новостным системам ИИ. На мой взгляд, эти системы весьма показательны с позиции изучения интеррогативной функции комплексного теста Тьюринга. Дело в том, что в работу новостных агентств средства ИИ стали внедряться давно, лет сорок назад. Важность и трудность организации правильных вопросов к новостным экспертным системам была прекрасно раскрыта Д.Деннетом в его работе «Эра интеллектуальных машин. Могут ли машины мыслить?» (1994 г.), которую он излагает на основе опыта работы с новостной системы CYRUS. Эта программа имитировала «знания» Сайруса Вэнса, секретаря в администрации американского президента Джеймса Картера (1976 – 1980 гг.). Компьютерная система много «знает» и способна проследить весьма широкий и глубокий цифровой след оригинала. Однако когда Д. Деннет задает банальный вопрос: «Может ли Сайрус завязывать шнурки», система «зависает». Дело в том, как он объяснил далее, что экспертные системы – это «деревни Потёмкина»: у них привлекательный интерфейс, однако внутри ничего нет, помимо «скелетных оснований слов».

Современные новостные системы, конечно, имеют более интересные программно-информационные основания. Но интересно было бы спросить у такой системы по аналогии с упомянутой новостной системой: А умеет ли Дмитрий Песков завязывать шнурки? У меня не получилось найти ответ на этот вопрос ни у одной современной поисковой web-системы.

Сегодня развивается своеобразная индустрия правильной постановки вопросов к системам ИИ. Недавно было проведение тьюринговое тестирование в Калифорнийском университете (Сан-Диего, США). Авторы работы воспользовались тьюринговой методологией и организовали 1400 игр, в которых 650 участ-

ников вели короткие беседы то с человеком, то с чат-GPT. Задача участников состояла в выявлении того, с кем они общаются – с компьютером или человеком. Оказалось, что 40 процентов участников ошиблись, посчитав, что общаются с человеком [9].

**3. Дефинитный вопрос общего ИИ** обеспечивает такое компьютерное определение когнитивного феномена, которое не вызывает когнитивного диссонанса. Например, совсем недавно, не более семи лет назад, «искусственный интеллект» как понятие – оксюморон у большинства гуманитарных специалистов, в частности, у философов вызывало вполне справедливое неприятие: интеллект может быть только у *homo sapiens* (!). Сегодня, в особенности с принятием указа № 490 [12] дела обстоят иначе. Например, не так давно, в октябре 2023 г. директор ИФ РАН и ведущий в мире специалист по этике Гусейнов А.А. заявил то, что сегодня человек вынужден налаживать отношения с искусственным интеллектом [10]. Вполне законен вопрос о том, с какими такими системами искусственного интеллекта я должен входить в связи и устанавливать всевозможные отношения. Носят ли это отношения интимный характер? Если да, то в какой степени? По всей видимости, выдающийся этик не способен оценить «этику искусственного интеллекта», слабо разбираясь в фактическом предмете искусственного интеллекта. Этика ИИ лежит в ином измерении, нежели чем «этика» со всевозможными ее прикладными вариациями. В свою очередь, этика ИИ опирается на концепцию общего ИИ, в которой много больше сверхчеловеческого, нежели чем человеческого. Здесь правомочен вопрос: Возможно ли научить общий искусственный интеллект этике?

К сожалению, научить этике искусственный интеллект нельзя. Обучить чему-то похожему на моральное поведение можно. Но научить конституировать правила этой моральной регуляции нельзя. Эту невозможность просто объясняется, помимо всего прочего, чисто формально-логическим путем. Дело в том, что машина работает исключительно с объемами понятий, экстенционально. С содержанием понятий, со смыслами, интенционально, работает человек. Человеку — интенциональное, компьютеру — экстенциональное. Именно таково фундаментальное отличие между человеком и системой ИИ. Экстенциональная работа не включает работу с идеальными объектами, точнее, исключает их. Этика – это теоретическая наука, она работает с идеальными сущностями, с интенционалами, а ими являются все без исключения понятия мировоззренческого плана: смысл, смерть, бессмертие, социальный идеал, свобода, совесть, вера, справедливость, правда и многое др. Этике искусственный интеллект на обучить.

Тем не менее совсем недавно в сфере этики ИИ насильственно смешалось интенциональное с экстенциональным. На международном форуме «Этика искусственного интеллекта: начало доверия» [9] рядом крупных организаций был подписан «Кодекс искусственного интеллекта», специально разработанный для регулирования морально-этической стороны использования технологии ИИ.

Документ подписали Правительство РФ и 20 российских компаний, среди которых — «Яндекс», «Сбер», VK, «Сколково», «Ростелеком», «МТС» и другие. Предполагается, что кодекс станет частью федерального проекта «Искусственный интеллект» и Стратегии развития информационного общества на 2017–2030 годы. Однако! В обсуждении и принятии этого документа нет ни одной организации, которая специализируется на проблемах этики. Почему отсутствуют философские структуры, в частности Институт философии РАН. Почему не привлекается НСММИ РАН, который на протяжении более 15 лет занимается специализацией проблем этики в контексте исследований искусственного интеллекта? Какое отношение к этике ИИ имеет артист Ф. Бондарчук, явившийся одним из главных подписантов Кодекса? [10]. Представим то, что автор настоящей статьи начнет сниматься в сериалах. Вполне объективно убеждение в том, что игра автора будет выглядеть нормально на фоне игры актеров большинства российских сериалов. Но вряд ли это будет соответствовать эстетическим требованиям. Подобного рода требования, скажем, когнитивной адекватности, нужно предъявить к подписантам кодекса этики в сфере ИИ. Мало кто понимает этику, совсем мало кто понимает ИИ, и, наконец, вряд ли кто-то понимает этике ИИ. Этикой должны заниматься этики. А подмена философов актерами, игнорирование этики, подмена этики другими регуляторами поведения – это прямая дорога нас всех в цифровое рабство. В этом рабстве окажутся и подписанты этого т.н. «Кодекса этики». Вряд ли «рейдерский» захват этики принесет государственную пользу.

**4. Конструкторский вопрос общего ИИ** раскрывает принципы работы компьютера, способного реализовать широкий спектр когнитивных феноменов. По причине интегральности проекта ОИИ, концепция вычислимости должна охватывать и принципы символизма и принципы коннекционизма. Символизм / коннекционизм представляет собой вычислительный аналог естественно-научного дуализма вещества/поля, дискретности/непрерывности, устойчивости/покоя и др. К сожалению, сегодня не предложены теоретические аналоги совмещения символьной и коннекционистской парадигм вычислимости, подобные формулам корпускулярно-волнового дуализма. Это заявка не будущее в сфере вычислительной математики. Именно здесь, на пути интеграции коннекционизма/символизма кроется загадка обнаружения «программистского Грааля» общего ИИ. На наш взгляд, такое совмещение связано с теоретико-алгоритмическими исследованиями по совмещению машины Корсакова и машины Тьюринга в едином формате [5]. В машине Корсакова отсутствует идея автоматных переходов, присущая специфике работы машины Тьюринга. В машине Тьюринга чрезвычайно примитивен способ представления лингвистических конструкций. Поэтому дуализм должен разрешаться на пути понимания того, что машина Корсакова – для коннекций, а машина Тьюринга – для инструкций. Совместно эти машины способны решить невероятно много вычислительных задач общего ИИ.

**5. Критический вопрос общего ИИ** отражает суть полемики по поводу возможности либо невозможности компьютерной реализации когнитивного феномена или класса когнитивных феноменов.

Оригинальный тест (А.Тьюринг,1950) отражает позитивные предположения о возможности компьютерной имитации диалогового интеллекта. А.Тьюринг предложил полемический стандарт (из девяти пропонирующих и оппонирующих положений, которые сегодня актуальны и представлены в многочисленных вопросах по поводу ИИ).

Тест на здравый смысл (Дж. Маккарти,1984, Д. Деннетт,1984) показывает ограниченность современных систем экспертного типа на предмет компьютерной имитации феномена «здравого смысла». Здравый смысл программно не реализуем. Разнятся подходы к трактовке «здравого смысла». Для Д.Деннета – это попытка представить информацию банального, глупого содержания. Для Дж. Маккарти, напротив, это попытка представить информацию, граничащую с интуитивными знаниями высокопрофессионального специалиста.

Китайская нация (Н. Блок,1978) изучает полемику по поводу правомочности компьютерной версии социологического реализма. Для естественных коммуникаций проблема неразрешима в вопросе: что первично – личность или общество? Для электронных коммуникаций мы полагаем то, что произошла трансформация социологической проблемы начала 20 века «М.Вебер / Э.Дюркгейм», которые отстаивали взгляды, соответственно, социологического номинализма / социологического реализма. В цифровой экономике вопрос звучит так: несколько реально функционально-компьютерное обобщение субъективных психических феноменов индивида на социальные системы. Функционализм, как правило, терпит поражение: роевое и глобальное сознание – миф, следовательно, ноосфера невозможна. Однако вполне убедительными являются противоположные доводы: когниции общества–машины реальны, но мы, люди, не способны оценить их реальность. Для этого требуется «расширенное сознание».

Китайская комната (Дж. Серль,1980) оценивает роль компьютерных средств в феномене «понимания». Понимание и вычисление – это совершенно разные стороны человеческой активности. Применение компьютеров отдаляет понимание. Поэтому компьютер понимать не может. Но часто он может весьма удачно имитировать способность понимания.

Тест зомби (Р. Кирк,1974, Д. Чалмерс,1995) включает большое разнообразие весьма эпатажных мысленных экспериментов с теми воображаемыми существами, которые не обладают когнитивными компетенциями, но весьма успешно имитируют обладание этими компетенциями. Для проекта искусственной личности это тест является самым фундаментальным и проявляется в следующем противоречии. По Д.Деннету: все мы, люди, - зомби, то есть неполноценные личности, так как, в частности, не обладаем полнотой восприятия и способностью реализовывать собственные мысли. Однако в условиях глобального ИИ

мы способны стать полноценными личностями. По Дж. Серлю всё выглядит с точностью до наоборот: все мы, люди, ранее (в золотом веке?) были полноценными личностями, в условиях же становления глобального ИИ мы превращаемся в зомби.

Параноидальный тест (К. Колби, 1980) раскрывает, что все традиционные компьютерные системы максимум на то, что способны – это имитировать параноидальное поведение, то есть поведение, ограниченное идеями программиста как источника навязчивой идеи, воплощаемого в алгоритме компьютерной программы.

Субкогнитивный тест (Р. Френч, 1990) показывает, что для полной аппроксимации человеческого когнитивного поведения компьютер (робот, аватар) должен жить как человек.

Гендерный тест (Ю. Генова, 1994) оценивает гендерные различия. Главное в тесте Тьюринга – не то, может ли машина мыслить, а то, чем мужчина отличается от женщины и, наоборот.

Инвертированный тест (С. Ватт, 1996) предлагает «обратить» тест. То есть судья должен быть судим и если наблюдатель присуждает ментальность игроку, то это – человек, так как компьютер воображать не может. То есть этот тест предлагает изучить машину Тьюринга, тестирующую тест Тьюринга.

Эмоциональный тест (А. Колмогоров, 1953; А. Сломан, 2000) оценивает феномен «любви» как основной ориентир эмоциональных машин. С функционалистской позиции любовь агента к машине возможна, если он не осознает того, что общается с машиной. Возможны различные нюансы.

Креативный тест (С. Брингсдйорд и др., 2000, В. И. Самохвалова, 2010, Пожарев, 2015) это главный тест на изучение принципов машинного творчества. Принципы оценки креативности включают теоретические наработки в исследованиях проблемы «творчество и машина» начиная с середины 19 века. К этим принципам относится следующее: аргумент Лавлейс (1843 г.): машина творить не может, по-настоящему творить может исключительно человек-программист; контраргумент Тьюринга (1950 г.): машина может творить, потому что компьютеры способны удивлять непредвиденным ходом выполнения программы; тест Лавлейс (2000 г.): машина не может творить, так как действует в условиях аксиоматической предзаданности; тест Лавлейс 2.0 (2014 г.): машина не может творить, но в условиях глубокого машинного обучения нужен специальный оценщик, чтобы отследить искусственность как бы естественных форм; тест Лавлейс 3.0 (Т. Пожарев с 2015 г. по сей день): оценщик должен владеть феноменологией смыслового восприятия артефакта, уметь работать со смыслом произведения.

Эти и другие частные тесты (подробнее см. в [9]) достаточно четко позволяют оценить возможности реализации критической функции комплексного теста Тьюринга. Имеется большое количество оснований для критики возможности реализации теста Тьюринга. Например, применительно к самому комплексному тесту возникают вопросы метауровневого характера: как алго-



ритмически представить критику критики. Вообще, что это такое? Очевидно, что подобного рода несуразный, точнее, гёделевский вопрос является типовым для системы общего ИИ.

**6. Организаторский вопрос общего ИИ** вырабатывает и поддерживает мероприятия, которые формируют комплексный ТТ как целостную систему планирования, подготовки и осуществления тестирования, а так же оценивания результатов его применения и распространения опыта. Среди организационных вопросов, пожалуй, определяющую роль играет спор о главных компетенциях специалистов, которые проводят тестирование. Эти компетенции определяются авторитетом специалистов в области средств ИТ – программных, информационных, технических, лингвистических, логических, математических, организационных, юридических и др. средств.

Проблема выделения главного в проекте общего ИИ – это важная задача управленческого характера. Главный специалист должен совмещать легальность / легитимность. Иначе когнитивная система «жить» не сможет. Выбор главного в ИТ – особая проблема, не имеет решения, ситуативно вариативна. Например, многие видные специалисты в сфере ИТ полагают, что основой любой компьютерной информационной системы служат лингвистические средства. То есть руководящими организационными силами в ИТ являются не программисты и датасайентисты, ни логики и математики, ни технари и юристы, а филологи, точнее, лингвисты. С этим частично можно согласиться. И не в связи с тем, что лингвисты отвечают за формализацию естественного языка, за унифицированные формы документов как основу человеко-машинного взаимодействия, за рекомендации со стандартами (напомню, например, про ключевую роль в проекте Интернета рекомендаций консорциума WWW). У лингвистов стоит более высокая задача, задача системного содержания. Они отвечают за формирование Единой системы классификации и кодирования информации (ЕСКК). В самом деле, если нет ЕСКК, нет и компьютерной системы как системы в целом. ЕСКК обеспечивает идентификацию, формализацию, систематизацию, дифференциацию, интеграцию, операционализацию и функционализацию всех подлежащих учету информационных элементов компьютерной системы. Если выявление нужности этих элементов возложить на лингвиста, то тогда лингвист в сфере ИИ будут выполнять роль аналитического философа, распутывая клубки терминологических противоречий и создавая благодаря этому четкую и ясную картину естественно-искусственного мира.

Главным в проекте общего ИИ является тот, кто способен строить проект общего ИИ исходя из позиции первенства своей собственной специализации. Лингвист должен считать то, что он главный в проекте ОИИ. Программист – то, что вокруг программиста должны свершаться все мероприятия. Информационщик – за хранение и обработку формализованной информации. Экономист должен просчитывать возможные пути эффективной реализации проекта. Логика и математики отвечают за универсальность и конкретность алгоритмов и



программ. Инженер – за работу техники. Юрист должен отвечать за защиту интеллектуальной собственности. Автор данной статьи выражает убежденность в том, что главным ответственным лицом в проекте общего ИИ должен быть философ – методолог техники.

**7. Конститутивный вопрос к общего ИИ** позиционирует отношение судьи (наблюдателя) к когнитивным аспектам компьютерной реальности. Комплексный тест Тьюринга способствует построению феноменологии компьютерного мира, обеспечивая постнеклассическое позиционирование человека как субъекта электронной культуры. Напомним, что классическое использование системы ИИ рассматривается по аналогии с исследованием объекта природы. Неклассическое позиционирование предполагает погружение наблюдателя-человека (Ч) в контекст многочисленных методов: каков метод – таков и предмет. Система ИИ функционирует в различных средах, конституированных различными идеологиями, т.е. политическими, экономическими, социальными и др. идеями, мифами, смыслами. Постнеклассическая парадигма утверждает о том, что что и человек и система ИИ проявляются в единстве естественно-искусственной системы как самоорганизующейся системы, непосредственно погруженной в смыслы общества и культуры).

- Классическая парадигма:  
Ч ↔ [ система ИИ = объект ]
- Неклассическая парадигма  
Ч ↔ [ Метод ↔ [система ИИ – среда]]
- Постнеклассическая парадигма  
[Ч ↔ [ Метод ↔ [система ИИ – социум ]]]

И самое главное, в рамках такого постнеклассического позиционирования проявляется уникальная роль человека в современном компьютерном мире. Функционализм любой разновидности ИИ – общего, сильного, слабого, глобального, доверенного и пр. – не способен компьютерно воспроизвести целостность, самость, яйность, смысл, творчество и мн.другое, составляющее суть homo sapiens, в силу определения понятия функции (функция как самоопределяющая рекурсия, функция как отображение и функция как роль). Других же методологических оснований ИИ просто-напросто нет, в силу определения ИИ. Поэтому самый общий вопрос общего ИИ – об уникальной роли человека – решается так: в условном противостоянии человека и системы общего искусственного интеллекта человек всегда главный. Угрозы и риски искусственного интеллекта – не более, чем метафоры, точнее, не более, чем значения функции как образы реальности. В реальности их нет. Реален человек, отчужденно производящий посредством ИИ эти угрозы и риски.

## Заключение.

Общие вопросы общего ИИ, поставленные в контексте исследования интегральных функций комплексного теста Тьюринга задают логически необходимые вопросы, без ответа на которые невозможен запуск проекта *общего искусственного интеллекта*. Предложенные вопросы проекта общего ИИ составляют основания для дискуссий по поводу возможностей, рисков и перспектив проекта общего ИИ в части его подпроектов: искусственной жизни, искусственного мозга/сознания, искусственной личности/зомби, искусственного общества, искусственного мира. Каждый подпроект общего ИИ с необходимостью упирается в требование концептуальной формулировки тех когнитивных функций, которые подлежат программно-аппаратной реализации. Конструктивная формулировка этих концептуальных средств представляется весьма эффективной, так как позволяет существенно экономить на разработке, внедрении и сопровождении компьютерных средств, инвариантных относительно разнообразия компонентов этих проектов. Либо экономить на закрытии тех проектов общего ИИ, которые, как показывает более скрупулезный анализ концептуальных вопросов, поставленных в этой статье, бесполезны или вообще бессмысленны.

## Ссылки

1. Алексеев А.Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. – М.: ИИнтелл, 2013 г. – 304 с.
2. Алексеев А.Ю. Общефункционалистский концепт искусственной потребности как основа общего искусственного интеллекта// Философские науки. Том: 62. № 11, 2019 г., С.111-124
3. Алексеев А.Ю. Объемная (3D) интенциональная семантика словаря искусственного общества // Ежеквартальный Интернет – журнал «Искусственные общества», Том 8, номер 1-4, I-IV квартал 2013 г. - С. 5-36
4. Алексеев А.Ю. Философия искусственного интеллекта в России с начала века по настоящее время// Наукovedческие исследования, 2022, № 1, С.96 - 120 // DOI: 10.31249/scis/2022.01.04
5. Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. Диссертации на соискание ученой степени доктора философских наук по специальности 09.00.08 – философия науки и техники, философский факультет МГУ имени М.В.Ломоносова, диссертационный совет Д 501.001.37, г. Москва, 2016 г.; URL: <https://istina.msu.ru/dissertations/15422339/>
6. Алексеев А.Ю. Формальное определение искусственного интеллекта. Доклад на семинаре «Мыслить вместе со сложностью», 12 июня 2022 г., ИФ РАН; URL: [https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IABl0o\\_XMiAB7vVeMn2xIs-EDvZ&index=77](https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IABl0o_XMiAB7vVeMn2xIs-EDvZ&index=77) (дата обращения: 1.10.2022).

7. Алексеев А.Ю, Янковская Е.А. Феноменология компьютерного творчества: тест Ады Лавлейс // Философия творчества: материалы Всероссийской научной конференции, 8-9 апреля 2015 г., Институт философии РАН, г. Москва. - М.: ИИнтелл, 2015. - С. 375-394.
8. В России при участии ВШЭ разработан и подписан Кодекс этики искусственного интеллекта. Национальный исследовательский университет Высшая школа экономики. Новости. URL: <https://www.hse.ru/news/522677221.html> (дата обращения: 11.10.2023).
9. ИИ может заменить людей в 40% случаев. [Электронный ресурс]. URL: <https://www.securitylab.ru/news/543329.php> (дата обращения: 11.10.2023)
10. Рабочий завтрак у Тосуняна, 14 октября 2023 г., г. Москва, Россия. URL: <https://scmai.ru/2023/10/14> (дата обращения: 11.10.2023).
11. Сайт международной конференции Artificial General Intelligence (AGI) [Электронный ресурс]. URL: <https://www.agi-conference.org/> (дата обращения: 11.10.2023)
12. Chalmers, D. Two-Dimensional Semantics (2006), "Two-Dimensional Semantics", in Oxford Handbook of Philosophy of Language, E. Lepore and B. Smith (eds.), Oxford: Oxford University Press, pp. 575–606, URL = <http://consc.net/papers/twodim.pdf> (дата обращения: 1.10.2022)
13. Pei Wang (ed.), Ben Goertzel, Stan Franklin. Artificial General Intelligence 2008: Proceedings of the First AGI Conference Series: Frontiers in Artificial Intelligence and Applications 171 IOS Press. 2008. - 508 p. - ISBN: 9781586038335; 1586038338
14. Shagrir O. "The Rise and Fall of Computational Functionalism", in Y. Ben-Menahem (ed.), Hilary Putnam, Cambridge: Cambridge University Press, 220–250 p.
15. Turing A.M. (1950). Computing Machinery and Intelligence, Mind 59 (236), pp. 433–460.

# GENERAL ISSUES OF ARTIFICIAL GENERAL INTELLIGENCE

A.Yu. Alekseev

**Abstract.** The project of artificial general intelligence (AGI), which implements a wide range of cognitive phenomena, can be conveniently studied on the basis of a study of the complex Turing test. This test is aimed at solving the main question of the philosophy of artificial intelligence “Can a computer do everything?”, where the universal quantifier runs through private Turing tests that answer the following questions: can a computer understand, live, create, be aware, love, be friends, etc.? Issues arising during the development of an AGI project can be conveniently studied by explicating the seven functions of a complex test: communicative, interrogative, organizational, definitive, critical and constructive and constitutive functions.

**Keywords:** artificial general intelligence, complex Turing test, formal definition of AI

**For citation:** Alekseev A.Yu. (2023). General issues of artificial general intelligence. *Journal of Digital Economy Research*, vol. 1, no 3, pp. 6–37. (in English). [DOI: 10.24833/14511791-2023-3-6-37](https://doi.org/10.24833/14511791-2023-3-6-37)

## ***Information about the author:***

**A.Yu. Alekseev,**

GAUGN, Faculty of Philosophy, 26 Maronovsky Pereulok, Moscow, 119049, Moscow  
RUDN, Engineering Academy, Department of Mechanisms and Control Processes, 115419, Moscow, Ordzhonikidze str. 3  
[aa65@list.ru](mailto:aa65@list.ru)

## Introduction.

### Main ideas of the project of artificial general intelligence

The phrase “artificial general intelligence” (AGI) can hardly be used as a term in national and departmental standards to implement relevant regulations. In recent years, this concept is often used in projects that have been given the proper name “GPT Chat” and associated with products of OpenAI, DeepMind, Anthropic and similar computer firms. However, AGI still belongs to the realm of science fiction to refer to technical systems capable of performing any cognitive task of an animal, human, or society. It is postulated that *general AI* systems are capable of realizing cognitive functions that belong not only to the sphere of *other consciousness*, but also to *other* and *alien consciousness*. That is, these functions are difficult to imagine. However, as suggested by AGI proponents, they can be imagined albeit in an intricate mathematical format. It is in this mathematical sense that attempts to rationally understand *general AI* are being made at the remarkable annual international AGI conferences (see the conference website [13] and the proceedings of the first 2008 conference [11]).

In this paper we will not search for a mathematical “Grail Cup” of *general AI*. This is the concern of the participants of AGI conferences: to find such a mathematical formula and implement a corresponding computer program that can work in a wide variety of cognitive fields. Our task is less ambitious: proceeding from a common position, relying on philosophical-methodological, holistic, conceptual, and semantic forms of analysis, to study the most general issues of *general AI*.

The first in a number of such questions is the definition of the concept. Under *artificial general intelligence* we will understand philosophical-methodological, scientific-theoretical and engineering-technological research focused on the construction and application of computer replications (imitations), representations (models) and reproductions (reproductions) of cognitive phenomena of a wide range of life, mental, personal and social manifestations.

General questions about the ways of development of this project are solved at the level of philosophy and methodology of artificial intelligence [5]. Earlier in [14] it was proposed to use the so-called “general functionalist approach” for conceptual organization of the *general AI* project. General functionalism continues the two main ideas of machine functionalism of H. Putnam. Putnam. The first idea is related to the functional homeomorphism of natural and artificial systems: the causal relations of brain and psyche are similar to the relations of a computing device to the logic of its functioning. The second idea - of multiple realization - asserts the invariance of cognitive function to the substrate of realization: the realization of the same function can be realized by the biological brain of a human being or by a computer system of artificial intelligence or by the tina of a Martian (as the brain of an agent possessing a *different* and *alien consciousness*).

Continuing classical machine functionalism, however, in our general version of functionalism we do not drive the human psyche into the determinants of the programmed logic of the universal human computing machine. This clearly anti-humanistic position was expressed by H. Putnam in the early, optimistic period of his work. Later, however, he made a radical skeptical turn: cognitive physics is as impossible as O. Comte's social physics [14]. In our opinion, such a turn should be softened. Indeed, the main thing in machine functionalism is the use of the idea of an indeterministic Turing machine (1936) as a technical means of realizing cognitive function. But H. Putnam used it incorrectly (in 1960): he did not take into account the fact that the Turing test (1950) does not identify the program with the way of its realization, but merely *evaluates* from the position of an observer (judge, another person) the *possibility* of invariance of functions, realizing them by means of a program of quasi-algorithmic format. The general functionalism of AI is test functionalism, based on the idea of testing rather than on the idea of identification.

The general functionalism of AI “stands on three whales”: the collecting, defining and observational functionalisms. *Collective functionalism* provides collection, identification, coordination, formalization, systematization, unification and codification of all kinds of functionalist theories. At the most general level, it fulfills the tasks of philosophical analysis (or, more precisely, analytic philosophy): unraveling the “tangle” of contradictions as a conflict of diversity of theoretical interpretations. In the specific conditions of “gathering” the standard bibliographic task of systematizing functionalist theories, empirical studies and their methodological conceptualizations is solved.

*Definitional functionalism* identifies the so-called “main” functionalist characteristics, relations, regularities, and causal connections. They are called main because the components of determinative functions are responsible for *essential, causally determined* realization, which ensures the achievement of goals under the condition of invariance of cognitive phenomena with respect to the structure and substrate of their realizer. For example, the main one is the relation of multiple realization of the phenomenon “normal physiological functioning of the human heart”, in which the heart works normally even in conditions of the corresponding pathology, because an artificial valve is responsible for pumping blood.

*Observational functionalism* evaluates from the position of an abstract or concrete observer (researcher, interrogator) various status of technological implementations and, in the course of conceptual generalization, identifies categorical parameters of the general AI project. Generalization implies categorization of the characteristics of general functionalism at the following levels, traditional for philosophical analysis: ontological, epistemological, logical, linguistic, axiological, aesthetic, ethical and praxeological. That is why, due to philosophical generalizations, this project is called “general AI”, and not because of the fact that mathematical manipulations allow us to apply the same software modules with the accuracy to the meaning of parameters, say, in the tasks of pattern recognition and in the tasks of formation of the movement trajectory of a walking robot.



## Composition of the overall AI project

Under such philosophical categorization, it is reasonable to divide the general AI project into five constituent parts - subprojects. Each modern subproject of AI research, as a rule, has transnational significance and is funded by a special class of cognitive phenomena (functions), the computer realization of which these projects implement. These are the following projects: artificial life, artificial consciousness (brain), artificial personality, artificial society.

The project of *artificial life* is focused on computer realization of vital phenomena and answers the questions: can a computer “live”? At least, about the plausibility of simulation by a hardware-software system of the phenomenon of origin, development, functioning and dying by an artificial system taking into account genetic mechanisms.

The tasks of computer realization of mental (psychical) phenomena form a project that sounds twofold, although it means the same thing: can a computer be conscious? Naturalistic trends call this project as “*artificial brain*”, metaphysical orientations designate this project by the phrase “*artificial consciousness*”.

The project “*artificial personality*” simulating personological phenomena (personality, individuality, freedom, self, meaning, creativity) was the first project in the field of AI. The fact is that the main work for the philosophy of artificial intelligence is A. M. Turing’s 1950 article. “Computing Machines and Intelligence” - demonstrated the capabilities of a universal digital computer and a trainable neural network (a child machine) to play the game of simulating various aspects of interpersonal communication.

The “*artificial society*” project programs social phenomena: can a computer be a society? This application of computer tools is necessary to model social behavior, more precisely, to study and reproduce the phenomena of intersubjective reality - friendship, love, hate, etc.

Finally, most recently, since October 2019, when Facebook Corporation was renamed Meta Corporation, the “*artificial world*” project has openly declared itself. “Can a computer do *everything*?” - is the main question of the new project. The concept of computability is involved in all the broadest, highest, and deepest, in all the lowest and shallowest manifestations of computer simulation of the world. The basic question of the Metaverse and the basic question of the Turing Comprehensive Test coincide.

## Basis of generalization of subprojects of general AI

It should be noted, however, that the subprojects outlined above, despite the diversity of cognitive phenomena, are based on the artificial intelligence project. At first, it is necessary to intellectually formalize the phenomenon that stands for “life”, “consciousness”, “personality”, etc., and only after such pre-project work is it possible to identify functional replications, representations and reproductions of a cognitive phenomenon with the phenomenon itself. For example, the basis of “artificial consciousness”

is the project of “artificial intelligence” in terms of programming representations of consciousness. There is no “artificial consciousness” as such in this subproject. There is only terminological confusion.

To avoid nominative difficulties, in [2] it was proposed to choose not the concept of “intelligence” (together with its correlates “data”, “knowledge” and “meanings”) and not “pain”, the concept of which is usually used by philosophers of “of *mind*” to demonstrate their own theoretical positions. The phenomenon of *need* turns out to be more productive. It provides a broad coverage of cognitive functions in the project of general AI, as it often serves as an external design of internal processes of thinking, understanding, realizing, loving, etc. The most important thing is that the concept of need is fixed. Most importantly, the concept of need captures the causal relationship between the “ought” and the “being”, what is *ought* and what is. Based on the functionalist modeling of need, the so-called “*artificial need*” emerges. It is this concept that, as it is proved in [2], is a candidate for the role of the main definition and conceptual basis for the integration of subprojects of general AI.

In this paper, we will not emphasize the methodology of the main cognitive phenomenon of the general AI project. It is possible that this phenomenon should be chosen not as “need” but as “creativity” (as it was suggested in [7]) and, after this choice, it would be advisable to work out subprojects in the style of Bergsonian evolutionary epistemology. It may well be that a class of cognitive phenomena needs to be considered in a generic AI project. Each of the phenomena is capable of playing the role of an integrator of subprojects of the general AI and depends on the specific preferences of the judge (observer) of the complex Turing test.

### Formal definition of artificial general intelligence

It is reasonable to consider the complex Turing test as a real, true, correct Turing test. In fact, often used in today’s scientific and near-scientific discussions, the notion of “Turing test” is not something concrete, empirically embodied. Turing Test (TT) is a collective notion for research of artificial intelligence system on the possibility of software and hardware realization of a wide variety of cognitive functions. The comprehensive TT consists of hundreds of versions of the original Turing “intelligence simulation game” (1950) and numerous versions of those versions. Today, three quarters of a century after the birth of the idea of the Turingian game of simulated intelligence, these private Turing tests reveal the rather complex cognitive phenomenology of the computer world of contemporary culture. Can a computer think, understand, create, be conscious, self-aware, love, be a person, etc.? In short, can a computer do *everything*? The quantifier of universality runs through the whole set of Turing tests, assessing the possibilities and prospects of hardware-software imitation of this very vast and multidimensional spectrum of cognitive phenomena. It is precisely such *universal* coverage that allows us to assert that a computer system realizing a comprehensive TT embodies the concept of *artificial general intelligence*.

The idea of a complex Turing test has been developed by the author for two decades. It is detailed, for example, in [5, 1]. In a recent paper [6], it was proved that the complex Turing test is a formal definition of AI, similarly to the fact that a Turing machine is a formal definition of the notion of “algorithm”. Indeed, the notion of a complex Turing test contributes to a compact and succinct formal definition of AI: *artificial intelligence is a quasi-algorithmic realization of a complex Turing test*. Perhaps the significance of the idea of a complex test is most emphasized by the distinction between logically sufficient and logically necessary conditions for AI development. Logically sufficient AI conditions are those that favor the development of AI. They include a hard-to-register set of directions, events, projects, etc. of activities widely presented in [4]. Logically sufficient conditions of AI development are many. They include essentially all those studies in AI methodology that at least do not harm AI development, but contribute to theoretical knowledge, engineering projects and attempts to link theory and practice in the field of AI.

The logically necessary conditions for development are much less. These are those conditions, in the absence of which AI development itself is impossible. The Turing test is such a necessary condition. The point is that the work on creation and development of AI systems should be preceded by the developer’s conviction that a computer (machine) can think, or will someday be able to think, or will never be able to think. The word “think”, as we noted above, represents and funds many other cognitive phenomena - understanding, consciousness, creativity, love... And what concerns these phenomena is within the competence of the Turing test. To avoid confusion with the notions of “AI” / “general AI” it is proposed to specify: not “Turing test”, but “complex Turing test”. Taking into account this clarification, it turns out that the *complex test sets the necessary conditions for the development of general AI, as it raises fundamental questions about the computer realization of a diverse range of cognitive phenomena*.

If it is found impossible to implement this or that private Turing test, i.e. if there is no positive decision on the conceptual realization of the chosen version of the private Turing test, it is fundamentally inexpedient to proceed to the development of a specific AI subproject embodying this private Turing test. And this is extremely important for the digital economy as a whole. If, for example, we prove the invalidity of the concept “cognitome”, it will significantly reduce the expenditure of resources on the dead-end branch of AI development in the part of the subproject “artificial brain”.

Proceeding from the importance of studying logically necessary conditions of AI, the existing complexity of the organization of the complex TT (a hundred of Turing tests!) and the hypercomplexity of the study of questions of meta-level order relative to the level of the study of sufficient conditions of AI, the relevance and importance of studying general questions about the development and implementation of the project of general AI on the basis of conceptual foundations provided by the complex Turing test is beyond doubt.

## Common functions of the complex Turing test in a generic AI project

The literature identifies at least a hundred major versions of the Turing test. We abstracted a number of versions earlier, generalized them in [5, 1] and proposed to study twenty-one versions as the most representative selection of so-called perfect private Turing tests. The ideas of perfect Turing tests are surjectively covered by A. Turing's concepts stated in the seminal work for artificial intelligence [15]. Let me remind you that in the original Turing test, although many words are used about the imitation of "mind", in fact the game of imitation of gender is revealed, performed by both a human and a computer as a digital "double" of a human.

For the purposes of our paper, this is not essential; nevertheless, the observation reveals a difficulty of formulation: what does true AI, as formulated in the Turing test (1950) outside of the various AI accidents as weak, strong, narrow, global, trusted, and finally general AI, actually do? It turns out that in the early days of AI, the subject area of "intelligence" lay in more general domains than "dialogical type intelligence". This again emphasizes the variability of the basic concept for general AI: need, creativity, gender. What other possible representations are possible?

It is convenient to organize the study of general issues of general AI by studying the functions of test functionalism common to Turing's private tests [1]. These functions of *generalization* (integration) of private tests include communicative, interrogative, definitive, critical, constructive, organizing, and constitutive functions. In this paper, these functions are interpreted in the format of questions that need to be solved in order to realize a comprehensive Turing test and thus to reveal the basic ideas of test functionalism of general AI. This is what will make it possible to explore the general issues of general AI.

**1. The communicative issue of generic AI** provides a preliminary interdisciplinary coordination of specialists who are involved in the creation of an AI project. The Turing test introduces metaphors, simple categories and distinct images that are understandable to physicists and lyricists, philosophers and programmers, linguists and lawyers, economists and chemists, and many others. To organize interdisciplinary interaction, it is reasonable to use the so-called 3D language of cognitive terms [3]. There are two interpretations of such 3D-language: naive and mature. The naive interpretation emphasizes the unity of three semantics of a cognitive term in accordance with the motives inspired by the so-called "philosophy of consciousness": mental, physical, and computational. The naive interpretation is elaborated by D. Chalmers in solving the psychophysical problem from the position of the so-called "naturalistic dualism", but with a very significant limitation - he ignored the computational context, which, in our opinion, distinguishes the third semantic metric. A more mature form of 2-dimensionality by D. Chalmers [12] defines a two-dimensional semantics based on the Kantian notions of a priori (first dimension) and a posteriori (second dimension). The first dimension as a rule covers the phenomena of consciousness, subjective reality. The second dimension is the naturalistic phenomena of objective

reality. But again D. Chalmers, even in his stricter interpretation of two-dimensionality, ignored the immersion of the researcher in the context of electronic culture, in the computational context. Again, the world's leading metaphysician lacks the concept of computability. We propose to orient ourselves to the realities of artificial intelligence and to functionalism as a native AI methodology [15]. That is, we have one dimension as *cognition*. These are a priori concepts, phenomena of consciousness. The second dimension is considered as *realization*, i.e. posterior concepts, phenomena that can be touched, felt, seen, heard, photographed - this is approximately how V.I. Lenin expressed it, characterizing the materialistic interpretation of reality. The third dimension is actually what constitutes the notion of "dimensions". Phenomena of the third dimension are represented as *computerizations* - computational algorithms, as well as the whole complex of program, information, technical, linguistic and other means of information technologies. Three-dimensional semantics expressing the semantic trinity of cognitive, technical and algorithmic is presented as a constructive basis for the language of communication, as metaphors, images, categories of general AI are strung on a sufficiently distinct conceptual framework.

**2. interrogative question of general AI** characterizes the Turing questions asked by the Turing judge (interrogator, observer) to the x-system about its y-capability from the position of the judge's own possession of z-competence. That is, this paper studies "the question of questions", or, more precisely, "the question of Turing questions". The importance of proper Turing questioning for an AI system has a long history, starting with A. Turing's 1950 paper. A little later it manifested itself in the study of questioning of AI news systems. In my opinion, these systems are quite indicative from the position of studying the interrogative function of Turing's complex test. The point is that AI tools started to be introduced into the work of news agencies a long time ago, about forty years ago. The importance and difficulty of organizing the right questions for news expert systems was perfectly revealed by D. Dennett in his work "The Age of Intelligent Machines. Can Machines Think?" (1994), which he describes on the basis of his experience with the CYRUS news system. This program simulated the "knowledge" of Cyrus Vance, a secretary in the administration of U.S. President James Carter (1976 - 1980). The computer system "knows" a lot and is able to trace a very broad and deep digital footprint of the original. However, when D. Dennett asks the trivial question, "Can Cyrus tie his shoelaces," the system "freezes." The point is, as he went on to explain, that expert systems are "Potemkin villages": they have an attractive interface, but nothing inside beyond "the skeletal bases of words".

Modern news systems, of course, have more interesting program-information bases. But it would be interesting to ask such a system by analogy with the mentioned news system: Does Dmitry Peskov know how to tie his shoelaces? I have not been able to find an answer to this question in any modern web search engine.

Today, a kind of industry is developing to correctly ask questions to AI systems. Recently, a turing test was conducted at the University of California (San Diego, USA). The authors of the work used the Turing methodology and organized 1400 games in



which 650 participants had short conversations with a human or a chat-GPT. The participants' task was to identify whether they were chatting with a computer or a human. It turned out that 40 percent of the participants were mistaken in thinking that they were communicating with a human [9].

**3. The definite question of general AI** provides such a computer definition of a cognitive phenomenon that does not cause cognitive dissonance. For example, not so long ago, not more than seven years ago, "artificial intelligence" as a concept - an oxymoron was quite rightly rejected by most humanitarian specialists, in particular, philosophers: only *homo sapiens* (!) can have intelligence. Today, especially with the adoption of Decree No. 490 [12], things are different. For example, not so long ago, in October 2023, A.A. Huseynov, the director of the Institute of Physics of the Russian Academy of Sciences and the world's leading expert on ethics, stated that today man is forced to establish relations with artificial intelligence [10]. It is quite legitimate to ask with what kind of artificial intelligence systems I should get in touch with and establish all kinds of relations. Are these relationships of an intimate nature? If so, to what extent? Apparently, an eminent ethicist is unable to appreciate the "ethics of artificial intelligence" with little understanding of the actual subject of artificial intelligence. AI ethics lies in a different dimension than "ethics" with all its various applied variations. In turn, AI ethics is based on the concept of general AI, which is more superhuman than human. Here the question is legitimate: Is it possible to teach artificial general intelligence ethics?

Unfortunately, you can't teach ethics to an artificial intelligence. It is possible to teach something similar to moral behavior. But it cannot be taught to constitute the rules of this moral regulation. This impossibility is simply explained, among other things, in a purely formal-logical way. The fact is that the machine works exclusively with volumes of concepts, extensionalistically. Man works with the content of concepts, with meanings, intensional. To the human being, intensional, to the computer, extensional. This is the fundamental difference between a human and an AI system. Extensional work does not include working with ideal objects, or rather, excludes them. Ethics is a theoretical science, it works with ideal entities, with intensifiers, and they are all the worldview concepts without exception: meaning, death, immortality, social ideal, freedom, conscience, faith, justice, truth, and many others. Ethics can be taught by artificial intelligence.

Nevertheless, very recently, the intensional has been forcibly mixed with the extensional in the field of AI ethics. At the international forum "Ethics of Artificial Intelligence: The Beginning of Trust" [9], a number of major organizations signed the "Code of Artificial Intelligence" specifically designed for AI ethics [9] a number of major organizations signed the "Code of Artificial Intelligence", specifically designed to regulate the moral and ethical side of the use of AI technology. The document was signed by the Russian Government and 20 Russian companies, including Yandex, Sber, VK, Skolkovo, Rostelecom, MTS and others. The code is expected to become part of the federal project "Artificial Intelligence" and the Information Society Development



Strategy for 2017-2030. However! In the discussion and adoption of this document, there is no organization that specializes in the problems of ethics. Why are there no philosophical structures, in particular the Institute of Philosophy of the Russian Academy of Sciences. Why is not involved NSMII RAS, which for more than 15 years has been specializing in the problems of ethics in the context of artificial intelligence research? What does the artist F. Bondarchuk, who was one of the main signatories of the Code, have to do with AI ethics? [10]. Let's imagine that the author of this article starts starring in TV series. It is quite objective belief that the author's game will look normal against the background of the game of actors of the majority of Russian serials. But it is unlikely that it will meet the aesthetic requirements. The same kind of requirements, say, cognitive adequacy, should be imposed on the signatories of the code of ethics in the field of AI. Few people understand ethics, very few people understand AI, and finally, hardly anyone understands AI ethics. Ethics should be handled by ethicists. And substituting philosophers for actors, ignoring ethics, and substituting ethics for other regulators of behavior is a direct road to digital slavery for all of us. The signatories of this so-called "Code of Ethics" will also find themselves in this slavery. It is unlikely that the "raider" seizure of ethics will be of any benefit to the state.

**4. The design issue of general AI** reveals the principles of a computer capable of realizing a wide range of cognitive phenomena. Because of the integrality of the OII project, the concept of computability should encompass both the principles of symbolism and connectionism. Symbolism/connectionism is a computational analog of the natural-scientific dualism of substance/field, discreteness/continuity, stability/rest, and others. Unfortunately, no theoretical analogs for combining symbolic and connectionist paradigms of computability similar to the formulas of the corpuscular-wave dualism have been proposed today. This is not a bid for the future in the field of computational mathematics. It is here, on the path of integration of connectionism/symbolism, that the mystery of discovering the "programmer's Grail" of general AI lies. In our view, this overlap is due to theoretical and algorithmic research on combining the Korsakov machine and the Turing machine in a unified format [5]. The Korsakov machine lacks the idea of automata transitions, which is inherent in the specifics of the Turing machine. The Turing machine has an extremely primitive way of representing linguistic constructs. Therefore, the dualism must be resolved by understanding that the Korsakoff machine is for connectives and the Turing machine is for instructions. Together, these machines can solve incredibly many computational problems of general AI.

**5. The critical question of general AI** reflects the essence of the controversy about the possibility or impossibility of computer realization of a cognitive phenomenon or a class of cognitive phenomena.

The original test (A. Turing, 1950) reflects positive assumptions about the possibility of computer simulation of dialogical intelligence. A. Turing proposed a polemical standard (of nine propositional and oppositional statements, which are relevant today and are presented in numerous questions about AI.

The common sense test (J. McCarthy, 1984, D. Dennett, 1984) shows the limitations of modern expert-type systems for computer simulation of the phenomenon of “common sense”. Common sense is not programmatically realizable. There are different approaches to the interpretation of “common sense”. For D. Dennett, it is an attempt to present information of banal, silly content. For J. McCarthy, on the contrary, it is an attempt to present information bordering on intuitive knowledge of a highly professional specialist.

The Chinese Nation (N. Block, 1978) examines the controversy over the validity of the computerized version of sociological realism. For natural communications, the problem is insoluble in the question: what is primary - the individual or society? For electronic communications we believe that there was a transformation of the sociological problem of the early 20th century “M. Weber / E. Durkheim”, who defended the views of sociological nominalism / sociological realism, respectively. In the digital economy, the question is: somewhat realistic functionalist-computer generalization of subjective mental phenomena of the individual to social systems. Functionalism is usually defeated: swarm and global consciousness is a myth, hence noosphere is impossible. However, the opposite arguments are quite convincing: the cognitions of the society-machine are real, but we, humans, are unable to assess their reality. This requires “expanded consciousness”.

China Room (J. Searle, 1980) assesses the role of computing in the phenomenon of “comprehension”. Comprehension and computation are quite different aspects of human activity. The use of computers alienates understanding. Therefore, a computer cannot understand. But it can often very successfully imitate the ability to understand.

The zombie test (R. Kirk, 1974, D. Chalmers, 1995) includes a wide variety of highly epathetic mental experiments with imaginary beings who do not possess cognitive competencies, but very successfully imitate the possession of these competencies. For the artificial personality project, this test is the most fundamental one and manifests itself in the following contradiction. According to D. Dennett: all of us humans are zombies, i.e. incomplete personalities, since, in particular, we do not possess the completeness of perception and the ability to realize our own thoughts. However, in the conditions of global AI we are capable of becoming full-fledged individuals. According to J. Searle, everything looks exactly the opposite: all of us, humans, used to be full-fledged individuals (in the golden age?), but in the conditions of global AI we are turning into zombies.

The paranoid test (C. Colby, 1980) reveals that all traditional computer systems are at most capable of doing is to simulate paranoid behavior, that is, behavior limited to the programmer's ideas as the source of the obsession embodied in the algorithm of the computer program.

The subcognitive test (R. French, 1990) shows that to fully approximate human cognitive behavior, a computer (robot, avatar) must live as a human.

The gender test (Yu. Genova, 1994) assesses gender differences. The main thing in the Turing test is not whether a machine can think, but how a man differs from a woman and vice versa.

The inverted test (S. Watt, 1996) proposes a “reverse” test. That is, the judge must be judged and if the observer awards mentality to the player, then it is human, since a computer cannot imagine. That is, this test proposes to study a Turing machine testing the Turing test.

The emotional test (A. Kolmogorov, 1953; A. Sloman, 2000) evaluates the phenomenon of “love” as the main reference point of emotional machines. From the functionalist position, an agent’s love for a machine is possible if he does not realize that he is communicating with a machine. Various nuances are possible.

The creativity test (S. Bringsjord et al., 2000, V.I. Samokhvalova, 2010, Pozharev, 2015) is the main test for studying the principles of machine creativity. The principles of creativity assessment include theoretical developments in the research of the problem “creativity and machine” since the middle of the 19th century. These principles include the following: Lovelace’s argument (1843): a machine cannot create, only a human programmer can truly create; Turing’s counterargument (1950): a machine can create because computers have ways of surprising unforeseen program execution; Lovelace’s test (2000): a machine cannot create because it operates under conditions of axiomatic predetermination; Lovelace’s test 2.0 (2014): a machine cannot create, but in the conditions of deep machine learning a special evaluator is needed to trace the artificiality of as if natural forms; Lovelace test 3.0 (T. Pozharev from 2015 to the present day): the evaluator must possess the phenomenology of the meaning perception of the artifact, be able to work with the meaning of the work.

These and other private tests (see [9] for details) allow us to evaluate quite clearly the possibilities of realizing the critical function of the complex Turing test. There are many grounds for criticizing the possibility of realizing the Turing test. For example, with respect to the complex test itself, there arise questions of meta-level character: how to algorithmically represent the criticism of criticism. In general, what is it? Obviously, this kind of absurd, or rather Gödelian, question is typical for a general AI system.

**6. Organizational issue of general AI** develops and supports activities that form a comprehensive TT as a holistic system of planning, preparation and implementation of testing, as well as evaluation of the results of its application and dissemination of experience. Among organizational issues, perhaps the dispute about the main competences of specialists who conduct testing plays a defining role. These competencies are determined by the authority of specialists in the field of IT tools - software, information, technical, linguistic, logical, mathematical, organizational, legal, etc. tools.

The problem of singling out the principal in a generic AI project is an important managerial task. The chief expert should combine legality / legitimacy. Otherwise, the cognitive system will not be able to “live”. The choice of the chief in IT is a special problem, has no solution, and is situationally variable. For example, many prominent

IT specialists believe that the basis of any computer information system is linguistic means. That is, the leading organizational forces in IT are not programmers and datascientists, nor logicians and mathematicians, nor technicians and lawyers, but philologists, more precisely, linguists. One can partially agree with this. And not due to the fact that linguists are responsible for formalization of natural language, for unified forms of documents as a basis for human-machine interaction, for recommendations with standards (I remind, for example, about the key role in the Internet project of the recommendations of the WWW consortium). Linguists have a higher task, the task of systemic content. They are responsible for the formation of the Unified Information Classification and Coding System (UICCS). Indeed, if there is no UCCS, there is no computer system as a whole. UCCS provides identification, formalization, systematization, differentiation, integration, operationalization and functionalization of all information elements of the computer system to be taken into account. If identification of the *necessity* of these elements is entrusted to a linguist, then the linguist in the field of AI will play the role of an analytical philosopher, unraveling the tangles of terminological contradictions and thus creating a clear and precise picture of the natural-art world.

The principal in a generic AI project is the one who is able to build a generic AI project from the position of primacy of his own specialization. A linguist should consider that he is the main person in the general AI project. The programmer - that all activities should be performed around the programmer. The information specialist is responsible for storage and processing of formalized information. Economist should calculate possible ways of effective realization of the project. Logicians and mathematicians are responsible for universality and concreteness of algorithms and programs. An engineer is responsible for the operation of machinery. A lawyer should be responsible for the protection of intellectual property. The author of this article expresses his conviction that the main responsible person in the project of general AI should be a philosopher - a methodologist of technology.

**7. The constitutive question to general AI** positions the attitude of the judge (observer) to the cognitive aspects of computer reality. The complex Turing test contributes to the construction of the phenomenology of the computer world, providing the postnonclassical positioning of man as a subject of electronic culture. Recall that the classical use of the AI system is considered by analogy with the study of an object of nature. Non-classical positioning implies immersing the human observer (H) in the context of multiple methods: what is the method is the subject. The AI system functions in different environments constituted by different ideologies, i.e. political, economic, social, etc. ideas, myths, meanings. The postnonclassical paradigm states that both the human being and the AI system manifest themselves in the unity of the natural-art system as a self-organizing system directly immersed in the meanings of society and culture).

- The classical paradigm:  
 $H \leftrightarrow [AI\ system = object]$
- Non-classical paradigm  
 $H \leftrightarrow \leftrightarrow [Method \leftrightarrow [AI\ system - environment]]$ .
- Postnon-classical paradigm  
 $[H \leftrightarrow \leftrightarrow [Method \leftrightarrow [AI\ system - society]]]$ .

And most importantly, within the framework of such postnonclassical positioning, the unique role of the human being in the modern computer world is manifested. The functionalism of any kind of AI - general, strong, weak, global, trusted, etc. - is not capable of computerized reproduction of integrity, selfhood, otiosity, meaning, creativity, and many others. - is not capable of computerized reproduction of the integrity, selfhood, ovality, meaning, creativity, and many other things that constitute the essence of homo sapiens, due to the definition of the concept of function (function as self-determining recursion, function as representation, and function as role). There is simply no other methodological basis for AI, by virtue of the definition of AI. Therefore, the most general question of general AI - about the unique role of man - is solved in the following way: in the conditional confrontation between man and the system of artificial general intelligence, man is always the main one. The threats and risks of artificial intelligence are nothing more than metaphors, or rather, nothing more than function values as images of reality. There is no such thing in reality. What is real is the human being who alienatedly produces these threats and risks through AI.

## Conclusion.

The general questions of general AI, posed in the context of the study of integral functions of the Turing complex test, pose logically necessary questions, without the answer to which it is impossible to launch the project of *artificial general intelligence*. The proposed questions of the general AI project form the basis for discussions on the possibilities, risks and prospects of the general AI project in terms of its subprojects: artificial life, artificial brain/consciousness, artificial personality/zombie, artificial society, artificial world. Each subproject of the general AI has to meet the requirement of conceptual formulation of those cognitive functions that are subject to hardware-software realization. Constructive formulation of these conceptual means seems to be very effective, as it allows to save on the development, implementation and maintenance of computer means invariant with respect to the diversity of components of these projects. Or to save on closing those general AI projects, which, as a more scrupulous analysis of the conceptual questions posed in this article shows, are useless or meaningless at all.

## References

1. Alekseev A.Yu. Complex Turing test: philosophical-methodological and sociocultural aspects. - Moscow: Intell, 2013. - 304 c.
2. Alekseev A.Y. General functionalist concept of artificial need as the basis of artificial general intelligence// Philosophical Sciences. Vol. 62. No. 11, 2019, P.111-124
3. Alekseev A.Yu. Volumetric (3D) intensional semantics of the artificial society vocabulary // Quarterly Internet Journal "Artificial Societies", Volume 8, Number 1-4, I-IV quarter 2013. - C. 5-36
4. Alekseev A.Yu. Philosophy of Artificial Intelligence in Russia from the beginning of the century to the present time// Science Studies, 2022, No. 1, P.96 - 120 // DOI: 10.31249/scis/2022.01.04
5. Alekseev A.Yu. Philosophy of artificial intelligence: conceptual status of the complex Turing test. Dissertation for the degree of Doctor of Philosophy in the specialty 09.00.08 - Philosophy of Science and Technology, Philosophy Department of Lomonosov Moscow State University, Dissertation Council D 501.001.37, Moscow, 2016; URL: <https://istina.msu.ru/dissertations/15422339/>.
6. Alekseev A.Yu. Formal definition of artificial intelligence. Report at the seminar "Thinking with Complexity", June 12, 2022, IF RAS; URL: [https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IABl0o\\_XMiAB7vVeMn2xIs-EDvZ&index=77](https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IABl0o_XMiAB7vVeMn2xIs-EDvZ&index=77) (accessed 1.10.2022).
7. Alexeev A.Y., Yankovskaya E.A. Phenomenology of computer creativity: the Ada Lovelace test // Philosophy of Creativity: Proceedings of the All-Russian Scientific Conference, April 8-9, 2015, Institute of Philosophy of the Russian Academy of Sciences, Moscow. - Moscow: Intell, 2015. - C. 375-394.
8. The Code of Ethics of Artificial Intelligence was developed and signed in Russia with the participation of HSE. National Research University Higher School of Economics. News. URL: <https://www.hse.ru/news/522677221.html> (accessed 11.10.2023).
9. AI can replace humans in 40% of cases. [Electronic resource]. URL: <https://www.securitylab.ru/news/543329.php> (accessed 11.10.2023)
10. Tosunyan's working breakfast, October 14, 2023, Moscow, Russia. URL: <https://sc-mai.ru/2023/10/14> (accessed 11.10.2023).
11. Site of the international conference Artificial General Intelligence (AGI) [Electronic resource]. URL: <https://www.agi-conference.org/> (accessed 11.10.2023).
12. Chalmers, D. Two-Dimensional Semantics (2006), "Two-Dimensional Semantics", in Oxford Handbook of Philosophy of Language, E. Lepore and B. Smith (eds.), Oxford: Oxford University Press, pp. 575-606. Smith (eds.), Oxford: Oxford University Press, pp. 575-606, URL = <http://consc.net/papers/twodim.pdf> (accessed 1.10.2022).
13. Pei Wang (ed.), Ben Goertzel, Stan Franklin. Artificial General Intelligence 2008: Proceedings of the First AGI Conference Series: Frontiers in Artificial Intelligence and Applications 171 IOS Press. 2008. - 508 p. - ISBN: 9781586038335; 1586038338
14. Shagrir O. "The Rise and Fall of Computational Functionalism," in Y. Ben-Menahem (ed.), Hilary Putnam, Cambridge: Cambridge University Press, 220-250 p.
15. Turing A.M. (1950). Computing Machinery and Intelligence, Mind 59 (236), pp. 433-460.