

**И. И. Абрамов**, д-р физ.-мат. наук, проф.,  
Белорусский государственный университет информатики и радиоэлектроники,  
Минск, 220013, Республика Беларусь, e-mail: nanodev@bsuir.edu.by

## ПЕРСПЕКТИВЫ И ПРОБЛЕМЫ СОЗДАНИЯ СВЕРХРАЗУМА. ЧАСТЬ II

Поступила в редакцию 08.11.2019

*Оценены перспективы и проблемы создания сверхразума. Методологической основой рассмотрения являются предложенные ранее полная электронная интерпретация функционирования мозга и теория сознания человека. В части II работы рассмотрены перспективы и проблемы создания сверхразума в рамках выделенного гибридного направления его разработки. Проанализированы угрозы для человечества, которые могут возникнуть в случае появления сверхразума.*

**Ключевые слова:** сверхразум, мозг, полная электронная интерпретация, сознание человека, наноэлектроника, математическое моделирование, искусственный интеллект

### Гибридное направление

В этом направлении используются компоненты как материалистического, так и идеалистического направлений. Проще говоря, реально оно, как правило, основано на применении электронного аппаратного обеспечения (*hardware*) и программного обеспечения (*software*). Очевидно, что здесь может быть большое число различных вариантов (сочетаний), поэтому далее кратко рассмотрим лишь обозначившиеся, основные. В последнее время также выделяется путь создания интерфейсов "человек — машина" или "человек — компьютер", который может быть отнесен к *wetware*\*. Нетрудно заметить, что идеалистическое направление [1], связанное с моделированием, также, строго говоря, относится к гибриднему направлению, так как требует использования для своей реализации компьютеров, т. е. материалистической составляющей.

Об особой перспективности гибридного направления свидетельствует хотя бы то, что в его рамках, как обосновано считает ряд специалистов, сверхразум уже создан — это Вы вместе с интернетом. Сюда же можно отнести и коллективный разум человечества и некоторые другие варианты коллективного разума. Однако, конеч-

но же, хотелось бы чтобы это были не некоторые распределенные системы, а все же реализация искусственного интеллекта человеческого уровня (ИИЧУ) и сверхразума была осуществлена в одном устройстве, желательно компактном, во всяком случае не "монстре" (см. [1]).

В настоящее время одним из наиболее перспективных считается метод обратного проектирования (разработки) мозга (не следует путать с полной эмуляцией мозга, см. [1]). В этом случае искусственный интеллект (ИИ) строится по образу человеческого мозга, как правило, с использованием как аппаратных компонентов, так и программного обеспечения. Иногда возможный результат этого пути образно называется "кремниевым интеллектом", "кремниевым сознанием", так как аппаратная часть реализуется на кремниевых ИС. Основная проблема здесь заключается в том, что в настоящее время мы не разобрались как следует в том, как функционирует мозг человека. Поэтому полная реализация этого метода (тут также может быть ряд вариантов), к сожалению, пока еще не достигнута. И здесь автор видит определенные перспективы для использования разработанных теории сознания и комплексного иерархического подхода исследования мозга [2—10] в качестве основы для построения таких систем.

\* Сленг *wetware* имеет много толкований. Наиболее простое и удачное — это смесь *software*, *hardware* и биологии.

К наиболее известным глобальным программам, в рамках которых предполагается использовать этот метод для построения подобных систем, следует отнести [11, 12]: BRAIN Initiative (США), Human Brain Project (НБР, ЕС). В рамках BRAIN Initiative это, прежде всего, проект коннектома человека [13], а в рамках НБР — нейрокибернетический подход, в котором будет осуществлено создание модели человеческого мозга на транзисторах (поведение нейронов имитируется с помощью транзисторов) с привлечением суперкомпьютеров [14]. Предполагается, что разработка будет вестись по пути от простого мозга к более сложному, а именно: мыши, крысы, кролика и кошки. Недостатки этих проектов рассматривались в работе [12].

О грандиозности задачи в этом случае свидетельствуют следующие данные. Так, в одном из проектов под руководством доктора Д. Модхи (IBM) с использованием компьютера Blue Gene компании IBM с производительностью более  $10^{15}$  флопс, удалось смоделировать всего лишь 4,5 % нейронов и синапсов человеческого мозга [11]. "Чтобы начать хотя бы частичное моделирование человеческого мозга, потребуется 880 000 процессоров... Он говорит об использовании не одного компьютера Blue Gene, а о тысячах таких компьютеров, которые займут не комнату, а целый городской квартал. Потребление энергии при этом было бы так велико, что для ее производства потребовалась бы атомная станция на 1000 МВт. А затем пришлось бы повернуть реку и пропустить ее сквозь схемы компьютера, чтобы отвести лишнее тепло и не дать им расплавиться" [11]. При этом замечу, что модель воспроизводит только таламокортикальную систему, а не весь мозг человека.

Как отмечалось в работе [12], столь же пессимистически настроен и другой руководитель проекта по методу обратного проектирования, в частности коннектома человека, профессор С. Сеунг. Его пессимизм в основном связан со сложностью завершения проекта ввиду грандиозного объема полной информации в коннектоме человека, оцениваемом в  $10^{20}$  байт.

Более оптимистически настроен руководитель другого совместного с IBM проекта, также реализующего метод обратного проектирования мозга в рамках НБР, — профессор Г. Маркрам (Швейцария). "Ключом к проекту Blue Brain доктора Маркрама служит так называемая "колонка неокортекса" — модуль, многократно повторяющийся в ткани мозга... Доктору Маркرامу потребовалось десять лет (с 1995 по 2005 г.), что-

бы составить карту нейронов в такой колонке и разобраться, как она работает. Когда расшифровка была закончена, он пришел в IBM, чтобы виртуально "размножить" полученную структуру... В 2009 на конференции TED он заявил, что может завершить проект за десять лет. (Скорее всего, он говорил об "урезанном" варианте человеческого мозга без всякой связи с другими отделами и органами чувств)" [11]. В своей работе он также использовал компьютер Blue Gene.

Убедительное и краткое резюме по этим работам сделано в монографии [11]: "Ученые, посвятившие свою жизнь обратной разработке мозга, понимают, что им предстоит десятки лет тяжелой работы. Но они убеждены, что их работа принесет практическую пользу. Они считают, что даже частичные ее результаты помогут раскрыть тайны психических расстройств, причиняющих людям страдания на протяжении всей истории человечества". И тем не менее. "Обратная разработка — это самое простое; после начинаются сложности — ведь нужно разобраться во всей этой информации" [11].

Перспективу здесь автор видит в разработке специальных методов (в том числе систем ИИ) автоматической иерархической обработки гигантских объемов информации. Это может быть, в частности, реализовано в рамках предложенного комплексного иерархического подхода исследования мозга, основанного на многоуровневом моделировании в сочетании с экспериментальными методами [2, 3, 6]. Другими словами, разработка методов анализа в целях имитации мозга и обработки информации должна осуществляться взаимосвязано, образно говоря, навстречу друг другу.

Среди других интересных исследований, прежде всего, отмечу SyNAPSE — программу DARPA (Управление перспективных исследовательских проектов Министерства обороны США) по финансированию создания систем нейроморфной адаптивной гибкомасштабируемой электроники по методу обратного проектирования мозга. Основная цель — создание электронных систем, имитирующих основные функции мозга млекопитающих. Участниками этой программы являются IBM, HRL Laboratories и ряд университетов.

По имеющейся информации [15] в результате выполнения проекта группой под руководством Д. Модхи создан "когнитивный компьютер" из тысяч соединенных параллельно кристаллов нанoeлектронных ИС, который имитирует работу 30 млрд нейронов и 100 трлн синапсов человеческого мозга, с быстродействием около  $10^{15}$  флопс.

"По утверждению IBM, интеллект SyNAPSE в значительной мере будет совершенствоваться на основе его взаимодействия с окружающим миром" [15].

Подобная система Neurogrid разработана группой специалистов в Стэнфордском университете. Это нейроморфная система для имитации нейронных модулей в реальном масштабе времени. Она позволяет моделировать миллион нейронов с миллионами синаптических связей, причем электронное устройство потребляет всего лишь 3 Вт энергии [16]. К другим подобным проектам можно отнести: SpiNNaker, HICANN. Замечу, что во всех этих четырех проектах используются различные архитектурные решения с применением отличающихся ИС.

Отмечу также другие финансируемые DARPA проекты — это CALO, NELL, а также в целом программа DARPA "Когнитивные вычислительные системы" с очень серьезными и амбициозными целями [15], ведущими, в сущности, в конечном итоге к созданию ИИЧУ.

*Несмотря на колоссальные сложности, автор считает метод обратного проектирования мозга перспективным направлением создания ИИЧУ.*

В то же время ряд специалистов (в основном компьютерщики) считает, что метод обратного проектирования мозга является затратным и можно попытаться создать более оптимальное устройство, чем мозг человека, т. е. без более или менее точного подражания ему. По мнению автора, провести подобную супероптимизацию, сделанную Природой, будет невероятно сложно, если ни невозможно [1]. Одним из таких проектов является Open Cog Б. Герцеля. В частности, предполагается, что когнитивная архитектура Open Cog Prime позволит создать ИИЧУ, и в конечном итоге — сверхразум. В основе концепции открытого проекта Open Cog лежит "то, что разум основан на высокоуровневом распознавании образов... Отдельные модули Open Cog будут выполнять такие функции, как восприятие, внимание и память. Делается это с помощью сложных, но индивидуально настроенных программных комплексов, включающих генетическое программирование и нейронные сети... Open Cog организует аппаратное и программное обеспечение в так называемую "когнитивную архитектуру", призванную моделировать деятельность мозга" [15].

Другая система LIDA Университета Мемфиса построена на основе теории глобального рабочего пространства, предложенной впервые Б. Баарсом [17], и, как предположили в журнале New Scientist, "демонстрирует признаки рудиментарного

сознания" [15]. "По когнитивной архитектуре LIDA напоминает Open Cog, а финансирует ее морское ведомство США" [15]. Замечу, что в проектах Open Cog и LIDA за основу все же взята определенная информация о мозге.

Не могу также не отметить ряд разработок компании IBM, которые напрямую не связаны с рассматриваемым вопросом, т. е. сильным ИИ, однако явно носят пионерский характер, а именно: Deep Blue, Watson. Deep Blue — это шахматный суперкомпьютер, который победил в матче 1997 года чемпиона мира Г. Каспарова. Основная задача суперкомпьютера Watson отвечать на вопросы, сформулированные устно человеком, используя доступ к 200 млн страниц информации, включая Википедию. Обработка информации идет со скоростью 500 Гбит/с с оперативной памятью 11 Тбайт. Дальнейшим развитием Watson стали программные продукты Watson Studio, Watson SDK. IBM Watson одержал яркую победу над двумя очень сильными игроками в телеигре Jeopardy, что многими считалось также невозможным. Несмотря на то что системы Deep Blue и Watson относятся к последнему поколению продвинутых экспертных систем (слабый ИИ), они являются важным опытом в создании мощных специализированных систем ИИ, и с точки зрения автора, подобные системы будут основным направлением развития ИИ ближайшего будущего.

Перспективным в рамках гибридного направления является создание разнообразных смесей (*wetware*) электронного аппаратного обеспечения (*hardware*), программного обеспечения (*software*) и биологии (в рассматриваемом случае различных систем человека, включая мозг). Число возможных вариантов (гипотетических и практически реализованных) просто огромно.

Прежде всего, отмечу пионерские опыты Х. Дельгадо с помощью "стимосиверов" — миниатюрных радиоэлектронных устройств "для передачи и приема радиосигналов, направленных к мозгу и от него" [18]. Эти опыты подтвердили "возможность установления прямой связи между мозгом и вычислительной машиной, минуя органы чувств, а также возможность автоматического обучения путем прямой передачи сигналов на определенные структуры мозга без участия сознания" [18]. Подчеркну, что работы были выполнены в середине прошлого века!

Отмечу также методы лечебной электрической стимуляции головного и спинного мозга, а также периферических нервов человека с применением "щадящих методов стимуляции", разрабо-

таннных под руководством академика Н. П. Бехтеревой, с помощью которых достигнуты значительные успехи в лечении самых различных заболеваний [19].

Успешными примерами интерфейсов "человек — машина" являются кохлеарные импланты для восстановления слуха, импланты сетчатки для восстановления зрения и т. п. При этом кроме различных нейроимплантов применяют и нейрочипы, т. е. специализированные ИС. Имеются успехи и в восстановлении двигательных функций (экзоскелеты), способности общаться (вспомним нейрокоммуникатор С. Хокинга) и других когнитивных функций [12]. Особый интерес представляют попытки создания протезов целых областей мозга, в частности гиппокампа, с помощью специализированных кремниевых ИС. Перспективным в этом случае может быть и использование различных микроэлектромеханических систем (МЭМС) и наноэлектромеханических систем (НЭМС) [12]. Краткое описание основных достижений в области интерфейсов "мозг — компьютер" дано в работе [11], а прогноз дальнейшего развития в работе [20].

В настоящее время предлагаются различные способы считывания мысли человека с применением искусственных устройств, включая нейрокомпьютерные интерфейсы, для передачи сообщений в мозг и связи с другим мозгом, вплоть до коммуникаций с интернетом путем создания Всемирной Сети Разума [2, 3, 7, 21], а это один из возможных путей к созданию сверхразума. Как уже отмечалось [2, 3, 7, 12], особую перспективу здесь представит использование достижений в области наноэлектроники, нанофотоники, наноматериалов, нанотехнологий и нанонаук в целом. Есть идея встроить оборудование смартфона непосредственно в тело человека [15]. Путь различных технологических дополнений человеческого тела, по существу, превращает человека в киборга, т. е. кибернетический организм. Р. Курцвейл считает, что киборгизация — это будущее человечества (см. далее).

В целом, *если и будут созданы действительно более компактные системы ИИЧУ и сверхразума по сравнению с отмеченными, то это, скорее всего, произойдет в рамках гибридного направления. В то же время можно условно выделить следующие основные направления развития ИИ ближайшего будущего: 1) продолжение усиления человеческого интеллекта; 2) симбиоз ИИ и человеческого интеллекта; 3) создание специализированных систем ИИ, особенно где необходимо обрабатывать большие массивы данных.*

## Угрозы

Не вызывает сомнения, что ИИ — научно-техническая область двойного назначения, а следовательно, обладает потенциалом угроз для человечества, причем в данном случае значительных.

Очень кратко и убедительно основные угрозы были выделены в блестящем учебнике по ИИ, а именно [22]: "...искусственный интеллект, по видимому, становится источником некоторых невиданных ранее проблем, в том числе перечисленных ниже... В результате автоматизации может увеличиться количество безработных. Может уменьшиться (или увеличиться) количество свободного времени, имеющегося в распоряжении людей. Люди могут потерять чувство собственной уникальности. Люди могут потерять некоторые из своих прав на личную жизнь. Использование систем искусственного интеллекта может привести к тому, что люди станут более безответственными. Успех искусственного интеллекта может стать началом конца человеческой расы". В этой же книге проведен краткий анализ каждой из сформулированных в мягкой форме проблем. Наиболее важная для нас — последняя. Сформулирую ее следующим образом: речь о будущем совместном существовании сверхразума и человечества. Остановимся на ней более подробно.

В целом, если предположить, что эволюция интеллекта будет происходить по "биологическому сценарию", а есть основания полагать, что так и будет, то возможны три основных варианта: 1) уничтожение менее развитого интеллекта более развитым; 2) подчинение менее развитого интеллекта более развитым; 3) мирное сосуществование интеллектов разного уровня. Поэтому не вызывает удивления то, что прогнозы специалистов по поводу нашего совместного существования со сверхразумом лежат в очень широком диапазоне, перекрывающем отмеченные варианты, от крайне пессимистического (уничтожение человечества) до эйфорически оптимистического (счастливого, безоблачного совместного существования).

**Пессимистические прогнозы.** В литературе описан ряд апокалипсических сценариев развития событий (см., например, [15, 23—25]), когда после возникновения сверхразума, как правило, качественного, в результате его действий уничтожается человечество. Так, целью книги [15] является: "Я хочу объяснить, почему катастрофический исход не просто возможен, но почти

неизбежен, если мы *сейчас* не начнем очень-очень тщательно к нему готовиться".

Анализ литературы показывает, что двумя главными проблемами, связанными со сверхразумом, являются: 1) будет практически невозможно предсказать поведение качественного сверхразума, если он появится; 2) в отличие от других технологий двойного назначения, для которых важно, чтобы они не попали в "злонамеренные руки", здесь, как справедливо отмечается в работе [22], "злонамеренные руки могут принадлежать самой технологии". Первое фактически означает, что нам необходимо обхитрить сверхразум, что, говоря образно, напоминает игру, когда ребенку необходимо по-настоящему переиграть взрослого. Возможно ли будет это? Второе же означает, что сверхразум может быть вне контроля человечества, т. е. ситуация меняется качественно! Анализ показывает, что ни один из методов контроля, самые разнообразные варианты которых подробно рассмотрены в работе [23], не обеспечивает полной надежной защиты. При этом анализировались методы из двух широких классов, а именно [23]: контроля возможностей сверхразума и его мотиваций. Поэтому *вывод автора — пессимистический сценарий развития событий, к сожалению, не является невероятным.*

Рассмотрим кратко ситуацию с безопасностью в области ИИ на настоящий момент времени, т. е., образно говоря, "спустимся на землю". К сожалению, она просто не вдохновляет. Известно огромное число случаев ошибок в работе систем ИИ, приводящих в том числе и к трагедиям, катастрофам. Наиболее известные и резонансные ошибки систем ИИ в исследованиях космоса, сферах производства, транспорта, энергетики, здравоохранения, финансов, связи, включая интернет, и других приведены в книге [25]. Среди последних трагедий достаточно вспомнить катастрофы двух самолетов Boeing 737 MAX 8, приведших к многочисленным человеческим жертвам, которые были связаны с ошибками в работе системы ИИ. Очень серьезную озабоченность вызывают многообразные хакерские (крэкерские) атаки, многочисленные известные примеры которых можно найти в интернете (достаточно упомянуть лишь известные масштабные утечки персональных данных). Особенно опасна автоматизация этих процессов. Большой вред может наносить и фейковое видео, основной целью которого является дискредитация людей, манипуляция общественным мнением. Обеспокоенность вызывает и все более широкое использование методик "черного ящика" при разработке

систем ИИ. В результате возникают очень серьезные вопросы с их безопасностью.

Однако наибольшую озабоченность вызывают на настоящий момент времени *автономные системы вооружения*, включая превращение дронов в снаряды. Нетрудно вообразить последствия сбоев в системах ИИ, управляющих оружием массового уничтожения, или крэкерских атак на них. Возможный итог — коллективный суицид человечества, т. е. омницид.

Таким образом, становится очевидным, что человечество пока не имеет значительных успехов (не может договориться) в преодолении менее существенных проблем по сравнению с теми, которые могут возникнуть в случае создания качественного сверхразума. Единственное, что утешает, так это то, что, если это и произойдет, то, по-видимому, это будет не скоро, т. е. у нас есть время и возможно немалое.

Еще более серьезным для человечества является *"суммирование рисков"*. О возможных проблемах для нас метко и емко высказался выдающийся английский ученый Стивен Хокинг [26]: "Планете сейчас грозит опасность в таком количестве областей, что мне трудно сохранить позитив. Опасности слишком велики и слишком многочисленны". К ним можно отнести: возможности глобальной экологической катастрофы, ядерной войны, других вариантов войн массового уничтожения, смертельной пандемии, столкновений с астероидами и другими массивными космическими объектами, истощение ресурсов Земли, ИИ, нано- и биотехнологии, генетика и др. Автор умышленно не отдает приоритет ни одному из рисков, так как, к сожалению, может случиться любой из них. Не исключено и их сочетание. "Суммирование рисков" попросту означает, что чем их больше, тем более вероятен катастрофический исход для человечества. И в то же время очевидно, что чудовищные средства тратятся не на то, на что следовало бы. В частности, триллионы долларов тратятся на вооружения, т. е. орудия уничтожения, а не для решения отмеченных проблем, угрожающих всему человечеству.

Итак, *главной проблемой в области ИИ является безопасность соответствующих систем.* С этой точкой зрения согласны многие специалисты в рассматриваемой области и те, кто просто интересуется данными вопросами. Результатом этой озабоченности явились "Асиломарские принципы разработки искусственного интеллекта" (см., например, [25]), количество подписей под которыми уже насчитывает тысячи извест-

ных мыслителей и разработчиков систем ИИ (см. ссылку <http://futureoflife.org/ai-principles>). Очень важно, чтобы разработка систем ИИ осуществлялась в соответствии с этими принципами. Несмотря на бесспорную важность этого добровольного документа, нарушение некоторых его параграфов может произойти неумышленно, или по неосторожности, или вследствие общего характера некоторых формулировок, носящих декларативный характер. Ведь очень сложный рассматриваемый вопрос. С точки зрения автора, необходимы более жесткие и контролируемые договоренности, в том числе на государственных и межгосударственных уровнях.

Основным направлением в области безопасности должно быть создание дружественного человеку ИИ, а в нашем случае — сверхразума. Как этого добиться? Пока, к сожалению, этого не знает никто!

Уже отмечалось, что различные методы контроля возможностей сверхразума вряд ли приведут к надежному полному контролю над ним. Большой успех может ожидаться от использования целого комплекса (иерархии) методов, вплоть до возможности его уничтожения. Очевидно, что последнее просто не гуманно относительно разума в любой его форме. И в то же время, учитывая многочисленные "грехи" человечества, в том числе и отмеченные выше, очень трудно предсказать как поведет себя сверхразум в отношении нас, если ни невозможно. Недаром Илон Маск в одном из ответов на вопрос в своем выступлении высказался очень эмоционально: "С искусственным интеллектом мы выпускаем дьявола!". Даже трудно вообразить, что может произойти, если "прирученный сверхразум" попадет в "злонамеренные" руки, а такое, к сожалению, не исключено.

**Оптимистические прогнозы.** ИИ уже давно существует по планете с огромным числом положительных практически важных результатов, прежде всего, в областях по обработке больших массивов данных, принятию решений, требующих оценки влияния множества факторов и возможных ситуаций. В подобных случаях человек, как правило, действует интуитивно и/или основываясь на предыдущем опыте. Нередко это оказывается малоэффективным и даже опасным. Таким образом, ИИ реально улучшает нашу жизнь. В связи с этим у многих специалистов есть очень веские основания для оптимизма в оценке будущего ИИ, т. е. он возник не на пустом месте.

Если говорить о сверхразуме, то главное свойство, которое должно быть заложено (вос-

питано) в нем, как считают многие, — это дружелюбие по отношению к человечеству. Однако возникает вопрос: как этого добиться? Ответ: путем правильного его воспитания. И здесь возникает другой вопрос: что означает "правильное воспитание", какую систему ценностей необходимо заложить в сверхразум? Череда вопросов может продолжаться и далее. И где в этом случае остановится?

Поэтому одно направление, которое фактически развивается уже давно, представляет, с точки зрения автора, перспективу — это симбиоз человека и систем ИИ. Ярким пропагандистом одного из вариантов этой идеи является Р. Курцвейл, который считает, что человечество сольется с машинами и в результате этого уцелеет [27]. Более того, проанализировав основные этапы развития человечества, известный нейробиолог и нейропсихолог Э. Голдберг пришел к выводу, что мы находимся "на пороге революции слияния" [28]. В этом случае, однако, человечество может столкнуться с рядом сложных и неожиданных проблем, связанных с совместимостью, а именно: информационными перегрузками; потерей своего личного "я" (идентичности); различными психическими расстройствами и др. Каковы будут последствия? Трудно предсказать.

Другая заслуживающая внимания идея, которая, на первый взгляд, кажется пессимистической, заключается в нашем качественном изменении оценки будущего ИИ, а именно: расценивать системы ИИ как "наших детей". Этой идеи придерживаются известные специалисты — Д. Хофштадтер и Х. Моравек. В частности, о роботах Х. Моравек написал: "Довольно скоро они станут способными вытеснить нас из жизни. Но я не очень встревожен такой возможностью, поскольку считаю, что будущие машины — это наше потомство, "порождение нашего разума", созданное по нашему образцу и подобию; это мы, но в более развитой форме. Как и биологические дети предыдущих поколений, они воплотят в себе лучшие надежды человечества на долгосрочное будущее... Дети нашего разума, не скованные тяжелой поступью биологической эволюции, будут вольны двигаться вперед и расти, чтобы в конце концов взять на себя безграничные и фундаментальные вызовы большой Вселенной... На какое-то время мы, люди, получим выгоду от их труда, но... как настоящие живые дети, они будут искать собственный путь и собственную судьбу, тогда как мы, их пожилые родители, молча отойдем в сторону".

## Заключение

Рассмотрение перспектив и проблем создания сверхразума проведено исходя из точки зрения, что человек — биологическая машина. Методологической основой при этом являлась полная электронная интерпретация функционирования мозга и теория сознания, предложенные автором ранее. На их основе выделены ключевые свойства мозга в качестве информационной системы, так как без понимания его работы трудно, а скорее всего невозможно, серьезно говорить об ИИЧУ, сверхразуме. Показано, что корректнее оценивать уровень интеграции мозга человека как объекта нанoeлектроники не по числу нейронов, а по числу каналов, которых гораздо больше —  $10^{19} \dots 10^{21}$ . Кроме фантастического уровня интеграции, для мозга выделены и ряд других важнейших свойств, которые пока не реализованы в ИС твердотельной электроники. Результатом этих особенностей мозга является сознание и квалиа, не достигнутые в современных системах ИИ. Сформулированный автором главный принцип функционирования мозга — принцип преобразований сигналов и энергетических реконструкций, с одной стороны, позволяет объяснить сознание и квалиа, а с другой стороны, — чем мозг качественно отличается от объектов твердотельной электроники. Отмечена важность аналого-цифровой обработки сигналов в мозге человека. Показано, что Природой была решена задача супероптимизации, которая еще не осуществлена в твердотельной электронике.

С этих позиций рассмотрены перспективы создания ИИЧУ и сверхразума. Выделены три направления их создания: 1) материалистическое; 2) идеалистическое; 3) гибридное. Проведен анализ всех трех направлений.

Показано, что в рамках материалистического направления ИИЧУ достижим. Оно же является вполне реальным и для достижения сверхразума. Сравнение с лучшими показателями твердотельной электроники, прежде всего, по уровню интеграции, энергопотреблению и габаритным размерам показывает, что мозг пока вне конкуренции. Так, только для достижения уровня интеграции мозга нам потребуется не менее 60 лет. И, тем не менее, приближенное воспроизведение сознания и квалиа в рамках электронного интеллекта, в принципе, возможно. Электронный интеллект перспективен для создания роботов с высоким уровнем ИИ, а также нейрокомпьютеров.

В то же время идеалистическое направление, по-видимому, будет наиболее безопасным, так

как ИИЧУ и сверхразум (в случае их создания) будут находиться не в реальном, а в виртуальном мире. По крайней мере, на начальном этапе. Отмечено, что приближенная и даже грубая полная эмуляция мозга может быть полезна в медицине, в частности, для "реставрации сознания" человека. По оценке автора, производительности самого мощного на сегодняшний день суперкомпьютера может быть недостаточно для детального моделирования (полной эмуляции) мозга человека. Отмечена проблематичность возникновения качественного сверхразума в результате "рекурсивного саморазвития". По мнению автора, и качественный сверхразум, если и появится, то все же будет допускать ошибки так же как и человек.

Особую перспективу для создания ИИЧУ и сверхразума представляет гибридное направление, сочетающее компоненты как материалистического, так и идеалистического направлений. Достаточно отметить, что коллективный сверхразум, как считают многие, уже создан — это Вы и интернет. Однако хотелось бы, чтобы ИИЧУ и сверхразум были не распределенными, а компактными устройствами. Рассмотрены ряд перспективных систем ИИ, которые могут считаться прообразами ИИЧУ.

И тем не менее, даже самые продвинутые (переводные) системы ИИ пока не могут сравниться с интеллектуальными возможностями человека по широте спектра и гибкости, способности к развитию. Они не обладают главным — сознанием и квалиа. Отмечу, однако, что появление качественного сверхразума будет одним из самых великих достижений человечества, так как он сможет взяться за решение грандиозных проблем, например, разработку методов лечения непобежденных болезней человека, распространение разума во Вселенной.

Рассмотрены угрозы для человечества, которые могут возникнуть с появлением качественного сверхразума. Проанализированы пессимистические и оптимистические прогнозы специалистов. К сожалению, самый печальный прогноз для будущего человечества не исключен. Поэтому основное внимание должно быть уделено безопасности.

Ближайшие проблемы для человечества, однако, связаны с автономными системами вооружений, использующими ИИ. Еще большие угрозы связаны с суммированием рисков.

В целом ИИ, образно говоря, — это "джин", которого человечество уже выпустило из бутылки. И этот процесс не подлежит возврату, так как человечество не может договориться по го-

раздо более простым вопросам и проблемам. ИИ будет развиваться и совершенствоваться далее. Из этого надо исходить как из непреложной истины. Главное при этом, чтобы "джин" был дружелюбен по отношению к человеку. Как это сделать? Сейчас этого не знает никто. И тем не менее, судя по всему, у нас будет время, чтобы разработать эффективные методы безопасности.

Перспективу в этом плане представляет идея слияния человека с машинами. Еще один взгляд импонирует автору — это то, что будущие системы ИИ — "наши дети". Этот взгляд полностью согласуется с вариантом "бессмертия", реализованного Природой, когда, образно говоря, мы передаем "эстафетную палочку" следующим поколениям. Главное чтобы "наши дети" уважали, помнили и любили своих родителей!

*Автор считает своим приятным долгом выразить искреннюю признательность моим ученицам Н. В. Колемейцевой, И. А. Романовой и И. Ю. Щербаковой за подготовку рукописи работы к печати.*

### Список литературы

1. **Абрамов И. И.** Перспективы и проблемы создания сверхразума. Часть I // Нано- и микросистемная техника. 2020. Т. 22, № 1. С. 46—56.
2. **Абрамов И. И.** Мозг как объект электроники. Saarbrücken: LAP LAMBERT Academic Publishing, 2012. 80 с.
3. **Abramov I. I.** Brain as an object of electronics. Saarbrücken: LAP LAMBERT Academic Publishing, 2013. 76 p.
4. **Абрамов И. И.** Мозг — объект органической гибридной наноэлектроники, или взгляд со стороны. Часть I // Нано- и микросистемная техника. 2013. № 1. С. 52—54.
5. **Абрамов И. И.** Мозг — объект органической гибридной наноэлектроники, или взгляд со стороны. Часть II // Нано- и микросистемная техника. 2013. № 3. С. 45—53.
6. **Абрамов И. И.** Мозг — объект органической гибридной наноэлектроники, или взгляд со стороны. Часть III // Нано- и микросистемная техника. 2013. № 5. С. 45—54.
7. **Абрамов И. И.** Мозг — объект органической гибридной наноэлектроники, или взгляд со стороны. Часть IV // Нано- и микросистемная техника. 2013. № 6. С. 49—53.
8. **Абрамов И. И.** Сознание человека, или возможности электроники. Часть I // Нано- и микросистемная техника. 2018. Т. 20, № 5. С. 308—320.

9. **Абрамов И. И.** Сознание человека, или возможности электроники. Часть II // Нано- и микросистемная техника. 2018. Т. 20, № 6. С. 368—384.
10. **Абрамов И. И.** Сознание человека, или возможности электроники. Часть III // Нано- и микросистемная техника. 2019. Т. 21, № 9. С. 555—574.
11. **Kaku M.** The future of the mind: the scientific quest to understand, enhance, and empower the mind. New York: Doubleday Publishers, 2014. 400 p.
12. **Абрамов И. И.** Перспективы использования наноэлектроники, наноматериалов и нанотехнологий в исследовании и медицине мозга человека // Нано- и микросистемная техника. 2016. № 1. С. 49—64.
13. **Seung S.** Connectome: How the brain's wiring makes us who we are. New York: Houghton Mifflin Harcourt. 2012. 384 p.
14. **Human Brain Project.** URL: <https://www.humanbrainproject.eu>
15. **Баррат Дж.** Последнее изобретение человечества: Искусственный интеллект и конец эры Homo sapiens. М.: Альпина нон-фикшн, 2019. 396 с.
16. **Neurogrid:** A mixed-analog-digital multichip system for large-scale neural simulations / B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alves — Icaza, J. V. Arthur, P. A. Merolla, K. Boahen // Proc. IEEE. 2014. Vol. 102, N. 5. P. 699—716.
17. **Cognition,** brain, and consciousness. Introduction to cognitive neuroscience. Second Edition / Ed. by B. J. Baars, N. M. Gage. Amsterdam: Elsevier, 2010. 658 p.
18. **Дельгадо Х.** Мозг и сознание. М.: Мир, 1971. 264 с.
19. **Лечебная** электрическая стимуляция мозга и нервов человека / Под общ. ред. Н. П. Бехтеревой. М.: АСТ; СПб.: Сова; Владимир: ВКТ, 2008. 464 с.
20. **Roadmap.** The future in brain/neural-computer interaction: HORIZON—2020. URL: <https://bnci-horizon2020.eu/roadmap>
21. **Хорост М.** Всемирный разум. М.: Эксмо, 2011. 288 с.
22. **Рассел С., Норвиг П.** Искусственный интеллект: современный подход, 2-е изд. М.: Изд. дом "Вильямс", 2006. 1408 с.
23. **Бостром Н.** Искусственный интеллект. Этапы. Угрозы. Стратегии. М.: Манн, Иванов и Фербер, 2016. 496 с.
24. **Что мы думаем о машинах, которые думают:** Ведущие мировые ученые об искусственном интеллекте / Под ред. Д. Брокман. М.: Альпина нон-фикшн, 2017. 549 с.
25. **Тегмарк М.** Жизнь 3.0. Быть человеком в эпоху искусственного интеллекта. М.: АСТ: CORPUS, 2019. 560 с.
26. **Хокинг С.** Краткие ответы на большие вопросы. М.: Эксмо, 2019. 256 с.
27. **Курвейл Р.** Эволюция разума. М.: Изд-во "Э", 2015. 352 с.
28. **Голдберг Э.** Креативный мозг: Как рождаются идеи, меняющие мир. М.: Эксмо, 2019. 384 с.
29. **Moravec H. P.** Robot: Mere machine to transcendent mind. Oxford University Press, 2000.

**I. I. Abramov,** Dr. Sci., Professor, e-mail: [nanodev@bsuir.edu.by](mailto:nanodev@bsuir.edu.by),  
Belarusian State University of Informatics and Radioelectronics, Minsk, 220013, Belarus

*Corresponding author:*

**Abramov Igor I.,** Dr. Sci., Professor, Belarusian State University of Informatics and Radioelectronics, Minsk, 220013, Belarus, e-mail: [nanodev@bsuir.edu.by](mailto:nanodev@bsuir.edu.by)

## Prospects and Problems of Supermind Development. Part II

*Received on November 08, 2019  
Accepted on November 15, 2019*

*The prospects and problems of supermind creation are estimated. The previously proposed full electronic interpretation of brain functioning and the theory of human consciousness are the methodological basis of consideration.*



In part II of the work the author examined the prospects and problems of artificial intelligence of the human level and supermind creation in the framework of the distinguished third area of their development — the hybrid development area. In this area, the components of both the materialistic and the idealistic development areas are used. This development area, apparently, is the most promising of the three distinguished areas of creation of artificial intelligence of the human level and supermind.

The threats to humanity that may arise if a high-quality supermind is created are analyzed. It is noted that both pessimistic and optimistic forecasts for the future of humanity can take place. In the near future, however, the greatest threats to humanity will be autonomous weapons systems, as well as the possible "summation of risks" from various threats.

**Keywords:** supermind, brain, full electronic interpretation, human consciousness, nanoelectronics, simulation, artificial intelligence

For citation:

**Abramov I. I.** Prospects and Problems of Supermind Development. Part II., *Nano- i mikrosistemnaya tekhnika*, 2020, vol. 22, no. 2, pp. 112–120.

DOI: 10.17587/nmst.22.112-120

## References

1. **Abramov I. I.** Prospects and problems of supermind development. Part I., *Nano- i mikrosistemnaya tekhnika*, 2020, vol. 22, no. 1, pp. 46–56.
2. **Abramov I. I.** Mozg kak obekt jelektroniki. Saarbrücken: LAP LAMBERT Academic Publishing, 2012. 80 p. (In Russian).
3. **Abramov I. I.** Brain as an object of electronics. Saarbrücken: LAP LAMBERT Academic Publishing, 2013. 76 p.
4. **Abramov I. I.** Mozg — obekt organicheskoy gibridnoj nanojelektroniki, ili vzgljad so storony. Chast' I, *Nano- i mikrosistemnaya tekhnika*, 2013, no. 1, pp. 52–54 (In Russian).
5. **Abramov I. I.** Mozg — obekt organicheskoy gibridnoj nanojelektroniki, ili vzgljad so storony. Chast' II, *Nano- i mikrosistemnaya tekhnika*. 2013, no. 3, pp. 45–53 (In Russian).
6. **Abramov I. I.** Mozg — obekt organicheskoy gibridnoj nanojelektroniki, ili vzgljad so storony. Chast' III, *Nano- i mikrosistemnaya tekhnika*. 2013, no. 5, pp. 45–54 (In Russian).
7. **Abramov I. I.** Mozg — obekt organicheskoy gibridnoj nanojelektroniki, ili vzgljad so storony. Chast' IV, *Nano- i mikrosistemnaya tekhnika*. 2013, no. 6, pp. 49–53 (In Russian).
8. **Abramov I. I.** Human consciousness, or possibilities of electronics. Part I. *Nano- i mikrosistemnaya tekhnika*, 2018, vol. 20, no. 5, pp. 308–320.
9. **Abramov I. I.** Human consciousness, or possibilities of electronics. Part II. *Nano- i mikrosistemnaya tekhnika*, 2018, vol. 20, no. 6, pp. 368–384.
10. **Abramov I. I.** Human consciousness, or possibilities of electronics. Part III. *Nano- i mikrosistemnaya tekhnika*, 2019, vol. 21, no. 9, pp. 555–574.
11. **Kaku M.** The future of the mind: the scientific quest to understand, enhance, and empower the mind. New York, Doubleday Publishers, 2014. 400 p.
12. **Abramov I. I.** Prospects of nanoelectronics, nanomaterials and nanotechnologies in research and medicine of the human brain, *Nano- i mikrosistemnaya tekhnika*, 2016, vol. 18, no. 1. P. 49–64.
13. **Seung S.** Connectome: How the brain's wiring makes us who we are. New York, Houghton Mifflin Harcourt, 2012. 384 p.
14. **Human Brain Project.** URL: <https://www.human-brainproject.eu>
15. **Barrat J.** Our final invention: Artificial intelligence and the end of the human era. New York, Thomas Dunne Books St. Martin's Press, 2013.
16. **Neurogrid:** A mixed-analog-digital multichip system for large-scale neural simulations / B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvares — Icaza, J. V. Arthur, P. A. Merolla, K. Boahen, *Proc. IEEE*, 2014, vol. 102, no. 5, pp. 699–716.
17. **Cognition, brain, and consciousness.** Introduction to cognitive neuroscience. Second Edition / Ed. by B. J. Baars, N. M. Gage. Amsterdam: Elsevier, 2010. 658 p.
18. **Delgado J. M. R.** Physical control of the Mind. Toward a psychocivilized society. New York, Harper & Row, Publishers, 1969.
19. **Medical electric stimulation of a brain and nerves of the human /** Pod obshh. red. N. P. Bekhterevoj. M.: AST; SPb.: Sova; Vladimir: VTK, 2008. 464 p. (In Russian).
20. **Roadmap.** The future in brain/neural-computer interaction: HORIZON—2020. URL: <https://bnci-horizon-2020.eu/roadmap>
21. **Chorost M.** World Wide Mind. Free Press, Simon & Schuster, 2011.
22. **Russel S. J., Norvig P.** Artificial intelligence: A modern approach. Second edition. Prentice Hall, 2003.
23. **Bostrom N.** Superintelligence. Path, dangers, strategies. N. Y., Oxford University Press, 2014.
24. **What to think about machines that think.** Today's leading thinkers on the age of machine intelligence, Ed. By J. Brockman, New York, Harper Perennial, 2015.
25. **Tegmark M.** Life 3.0: Being human in the age of artificial intelligence. Vintage Books, 2017.
26. **Hawking S.** Brief answers to the big questions. Space-time Publications Ltd., 2018.
27. **Kurzweil R.** How to create a mind: The secret of human thought revealed. Viking, 2012.
28. **Goldberg E.** Creativity: The human brain in the age of innovation, Oxford University Press, 2018.
29. **Moravec H. P.** Robot: Mere machine to transcendent mind, Oxford University Press, 2000.

Адрес редакции журнала: 107076, Москва, Строминский пер., 4. Телефон редакции журнала (499) 269-5510. E-mail: [nmst@novtex.ru](mailto:nmst@novtex.ru)  
Журнал зарегистрирован в Федеральной службе по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия.  
Свидетельство о регистрации ПИ № 77-18289 от 06.09.04.

Технический редактор Т. А. Шацкая. Корректор Е. В. Комиссарова.

Сдано в набор 17.12.2019. Подписано в печать 22.01.2020. Формат 60×88 1/8. Заказ МС220. Цена договорная  
Оригинал-макет ООО «Авансед солошнз». Отпечатано в ООО «Авансед солошнз». 119071, г. Москва, Ленинский пр-т, д. 19, стр. 1. Сайт: [www.aov.ru](http://www.aov.ru)