

Эгоистичный искусственный интеллект¹

О. А. Лукьянова, О. Ю. Никитин

Вычислительный Центр ДВО РАН
680000, Хабаровск, ул. Ким Ю Чена, 65

e-mail: ollukyan@gmail.com, olegioner@gmail.com

Аннотация. В статье предлагается полный минимальный список свойств, присущих интеллектуальным системам на человеческом уровне. Интеллект представлен с алгоритмической точки зрения и рассматривается как способность генерировать принципиально новые алгоритмы сжатия. Сознание рассмотрено как важное свойство интеллектуальной системы. Предложен подход к сознанию как механизму интеграции информации из трех компонентов. Показано, что интеллектуальная система должна состоять из отдельных самоорганизующихся частей. Предложено понятие эгоистичного искусственного интеллекта как подхода к созданию интеллекта, достигающего человеческого уровня, основанного на принципе саморазвития, автономности и самоорганизации, а не достижению каких-либо целей, поставленных перед интеллектуальной системой. Предложено четыре постулата интеллекта, которые могут помочь прояснить понимание сильного искусственного интеллекта.

Ключевые слова: эгоистичный сильный искусственный интеллект, сильный искусственный интеллект, колмогоровская сложность, самоорганизация, алгоритм сжатия, агент.

1. Введение

Искусственный интеллект в узком и широком смысле обычно определяется как способность агента достичь определенных целей. Различают сильный и слабый искусственный интеллект. Разница между этими понятиями заключается в степени универсальности целей, которые преследует агент.

Естественный интеллект обычно рассматривается как стандарт для сильного искусственного интеллекта. Тем не менее сильный искусственный интеллект может быть представлен биологической системой, которая является результатом эволюции, искусственной системой, спроектированной исследователем или даже случайным возникшим агентом во вселенной (больцмановский мозг) [1].

Для разработки и оценки абстрактных моделей сильного интеллекта необходимо сформулировать полный минимальный список свойств, присущих системам с интеллектом человеческого уровня, независимо от природы происхождения такой

¹ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-31-00188. Некоторые из подходов и технологий, используемых в работе, были разработаны в рамках выполнения работ по исследовательскому проекту РФФИ № 18-29-03196.

системы. В этой статье приводится ряд постулатов и теоретических обоснований концепций интеллекта и сознания, предназначенных для получения единой методологической оценки и построения систем сильного искусственного интеллекта на основе автономии агента.

2. Подходы к определению интеллекта

Поскольку статья посвящена получению унифицированных критериев интеллекта, ниже дается обзор теорий интеллекта, подходящих для его универсальной (не ориентированной на человека) формулировки. Например, Колер и Вертхаймер [2, 3] считали интеллект способностью перестраивать концепцию исходной проблемной ситуации (гештальт-психологическая теория интеллекта). По их мнению, возникновение решения связано с реструктуризацией поля восприятия агента, в котором распознаются отношения между элементами, необходимыми для решения проблемы.

Онтологическая теория интеллекта [4] изображает интеллект как особую форму сознания. У агента есть определенный набор знаний о мире (декларативные знания) и способы взаимодействия с окружающей средой (процедурные знания). Агент определяется характеристиками этой базы знаний.

Гештальт-психологическая теория интеллекта не учитывает роль прошлого опыта, тогда как онтологическая теория интеллекта, напротив, считает эффективную организацию прошлого опыта необходимым условием эффективной интеллектуальной работы. Обе модели рассматривают интеллект с точки зрения содержания.

Согласно операциональной теории интеллекта [5] интеллект — это самая совершенная форма доступа к окружающей среде, представляющая единство процесса ассимиляции (решение проблемной ситуации происходит с помощью имеющихся схем действия или когнитивных схем) и процесса аккомодации (изменение этих когнитивных схем в зависимости от требования объективного мирового взаимодействия). Таким образом, суть интеллекта заключается в том, чтобы поддерживать гибкую и одновременно устойчивую адаптацию к физической и социальной реальности, а его основная цель — структурировать (организовать) взаимодействие человека с окружающей средой.

Согласно теории умения [6] интеллект подразумевает наличие умственных способностей, таких как «способность обосновывать, планировать, решать проблемы, мыслить абстрактно, сравнивать сложные идеи, быстро учиться, использовать накопленный опыт». В области сильного искусственного интеллекта придерживаются такого подхода, а интеллект определяется как «способность достичь сложных целей в сложной среде» [7]. Этот подход из-за его концентрации на достижении

определенных целей, используя сложные когнитивные инструменты, называется целеустремленным интеллектом (Goal-Oriented Intelligence (GOI)) [8]. Существует описание ограничений этого подхода, связанных с необходимостью внешнего аппарата для передачи целей. Авторы предлагают новый подход, основанный на неограниченном прогрессивном развитии [Open-Ended Intelligence (OEI)] в метастабильном состоянии [8]. Этот подход уделяет особое внимание различию между агентом и окружающей средой.

Идеи эмерджентности интеллекта в условиях автономии агентов рассматриваются в теории автономного адаптивного управления (ААУ) [9]. Интеллект в данной теории — это способность автономного агента принимать решения, которые согласуются с целями автономного управления на основе оценочной базы знаний и системы оценки состояния объекта, которые накапливаются в процессе жизни агента, и в дальнейшем используются. В то же время цели автономного управления рассматриваются как эмерджентные, т. е., обладающие свойствами, возникающими в результате объединения элементов системы. Они основаны на двух основных приоритетах: выживание объекта управления и накопление знаний в базе знаний.

В теории ААУ рассматриваются четыре основных условия интеллекта: «условие автономности» — агент явно отделен от окружающей среды и принимает решения полностью автономно с момента возникновения; «условие дискретности» — конечный набор элементов системы, дискретных во времени; «условие максимальной начальной приспособленности» — наследование, позволяющее агенту действовать с некоторым начальным достаточным и эффективным уровнем приспособленности; и «условие минимума исходных знаний» — база знаний агента должна быть первоначально заполнена только до уровня поддержания начальной приспособленности. Существует связь между принципами максимальной начальной приспособленности и минимумом исходных знаний. Эти основные условия исчерпывающе описывают понятие автономии.

Таким образом, подход автономного искусственного интеллекта предполагает рассмотрение взаимосвязи между автономным объектом управления, системой управления, находящейся внутри объекта управления, и внешней средой. В подходе автономного адаптивного управления отсутствуют сложные когнитивные инструменты, специфичные для подхода GOI, а также механизмы их возникновения, описанные в подходе OEI.

В работе [10] описано применение подхода минимальной длины сообщения (MML) как механизма функционирования интеллекта с использованием алгоритмов сжатия. Двухкомпонентная количественная оценка эффективности интеллекта в отношении сложности входных данных и их описания позволяет понять уровень интеллекта агента.

Тем не менее такая оценка не указывает на создание принципиально новых алгоритмов сжатия в закрытой информационной системе «агент — окружающая среда». Все интеллектуальные алгоритмы обработки данных в соответствии с этим подходом являются продуктами знаний, уже находящимися в системе «агент — окружающая среда». В связи с этим возникает вопрос о механизме генерации абсолютно новых знаний. Согласно методу MML все знания человечества — это только рекомбинация тех, которые существовали ранее (включая наблюдаемые в природе).

Ниже предложен теоретический подход, который должен позволить оценивать автономные интеллектуальные системы независимо от их происхождения (биологического или искусственного), а также строить новые и объяснять процедуру появления новых знаний с использованием случайных процессов.

3. Постулаты эгоистичного сильного искусственного интеллекта

Результаты анализа предыдущих исследований концепции интеллекта человеческого уровня как абстрактного явления, показали объективную необходимость выполнения определенных требований к интеллектуальным системам. Так, авторы статьи предлагают список критериев интеллекта:

- интеллект предполагает создание принципиально новых алгоритмов сжатия и имеет стохастическую составляющую;
- интеллектуальная система обладает сознанием, необходимым для достижения автономии, интегрирующим информацию о себе, окружающей среде и своем месте в окружающей среде;
- интеллект имеет аппарат для создания сложных абстрактных моделей;
- интеллект характеризуется эмерджентной самоорганизацией.

На рис. 1 показана схема закрытой информационной системы «агент — окружающая среда», отражающая критерии 1–4. Агент имеет базу данных алгоритмов (AB), содержащую набор алгоритмов, используемых для обработки информации, а также генератор алгоритмов (AG), содержащий алгоритмы сжатия. Здесь агент явно отделен от окружающей среды, является автономным и действует, исходя из собственных интересов. Агент объединяет информацию о себе и окружающей среде, имеет аппарат для генерации новых алгоритмов (сложных абстрактных моделей) и создает образы на основе базы данных алгоритмов и сенсорной информации из среды. Система управления регулирует состояние контролируемого агента в соответствии с характеристиками состояния тела агента (BS). Окружающая среда также имеет стохастический генератор, который осуществляет стохастическое смешивание и обогащение информации и способствует самоорганизации принципиально новых алгоритмов.

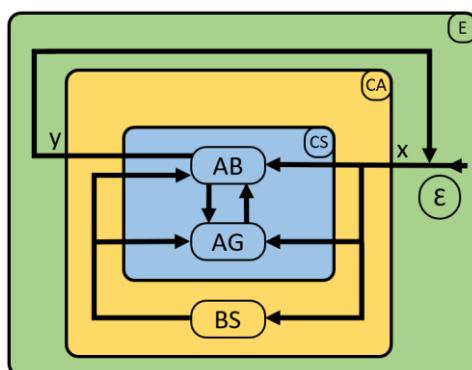


Рисунок 1. Схема закрытой информационной системы «агент — окружающая среда». E — среда; CA — управляемый агент; CS — система управления; AB — база данных алгоритмов; AG — генератор алгоритмов; BS — состояние тела; x — входные данные; y — выходные данные; \mathcal{E} — стохастическая переменная

Система, имеющая сознание, действующая автономно и руководствующаяся собственными интересами, может называться «эгоистичной». В соответствии со списком вышеприведенных критериев авторы предлагают определение интеллекта как автономной системы поддержания метастабильности в сложной среде, действующей в соответствии с ее интересами, обладающей сознанием, а также способной создавать и улучшать адаптивные алгоритмы своей работы. Это определение можно назвать «эгоистичным сильным искусственным интеллектом» и рассматривать как возможный подход к оценке и развитию интеллекта человеческого уровня.

Ниже рассмотрим каждый из вышеперечисленных критериев интеллекта более подробно.

3.1. Интеллект предполагает создание принципиально новых алгоритмов сжатия и имеет стохастическую составляющую

В этом параграфе будет показано, что создание принципиально новых знаний требует генерации алгоритмов с коэффициентом сжатия, большим, чем у предыдущих. Кроме того, применяя свойства колмогоровской сложности, будет доказано, что в замкнутой информационной системе для этого требуется стохастическая составляющая.

Определение 1. Пусть фиксирован некоторый способ описания (декомпрессора) D . Для данного слова x рассмотрим все его описания, т. е. все слова y , для которых $D(y)$ определено и равно x . Длина кратчайшего из них называется колмогоровской сложностью слова x для метода описания D :

$$KS_D(x) = \min \{l(y) \mid D(y) = x\}.$$

Здесь и далее $l(y)$ обозначает длину слова y . Индекс D подчеркивает, что определение зависит от выбора метода D [11].

Представьте себе агента с врожденным (унаследованным) набором алгоритмов выживания в сложной динамической среде, наделенного сознанием и сложными познавательными инструментами. Можно ли назвать этого агента по-настоящему интеллектуальным, если у него нет аппарата для создания принципиально новых алгоритмов действия? Предположим, что врожденный набор алгоритмов является исчерпывающим для данной среды. Таким образом, агент сможет достичь своих целей, как высших (выживание и накопление знаний), так и динамических (связанных с конкретными ситуациями). Если это так, то весь репертуар возможных действий агента будет предопределен. Даже если агент оснащен возможностью генерации новых алгоритмов действия, такой агент не сможет создавать принципиально новые алгоритмы. Здесь под принципиально новыми алгоритмами авторы имеют в виду алгоритмы с меньшей колмогоровской сложностью (KS), чем у предыдущих алгоритмов, применявшихся для решения конкретной задачи. Тогда получаем агента — интеллектуального адаптивного исполнителя генетической программы, который имеет сознание, но не способен к высокоуровневому творчеству. Поскольку человек определенно является системой, способной генерировать принципиально новые алгоритмы, авторы предлагают включить способность высокоуровневого творчества в определение сильного (человеческого) интеллекта.

Интеллект часто представляется как процедура сжатия [10]. Авторы предлагают подход к определению интеллекта, основанный на колмогоровской сложности, согласно которому *интеллект представляет собой процесс минимально сложного описания различных явлений*.

Определение 2. Пусть алгоритм A преобразует информацию в слове x в слово y . Создание принципиально новых алгоритмов — это минимизация колмогоровской сложности алгоритма A , содержащего максимальный объем информации о слове x .

$$\begin{cases} KS(A) \rightarrow \min; \\ I(A : x) = KS(x) - KS(x | A) \rightarrow \max, \end{cases}$$

где $KS(A)$ — колмогоровская сложность алгоритма A ; $I(A : x)$ — информация о слове x в алгоритме A ; $KS(x)$ — колмогоровская сложность слова x ; $KS(x / A)$ — колмогоровская сложность слова x в алгоритме A .

Результатом работы алгоритма A для данных x является y :

$$A(x) = y.$$

Алгоритм A сжимается с использованием базового алгоритма O :

$$O(A) = A'.$$

По определению 2 создание принципиально нового алгоритма сжатия входной информации невозможно без создания алгоритма с колмогоровской сложностью ниже, чем у предыдущего алгоритма.

Алгоритм O создает алгоритм A_1 такой, что его колмогоровская сложность меньше, чем колмогоровская сложность алгоритма A_0 для одного и того же набора данных.

$$KS(A_1) < KS(A_0).$$

Алгоритм O имеет определенную максимальную степень сжатия. Таким образом, согласно аксиоме невозрастающей колмогоровской сложности [11] колмогоровская сложность алгоритма A больше или равна колмогоровской сложности результата сжатия алгоритма A с использованием алгоритма O :

$$KS_o(O(A)) \leq KS_o(A) + c.$$

Из двух предыдущих уравнений получаем процесс преобразования алгоритма A_0 в A_1 с использованием O

$$KS_o(A_1) < KS_o(O(A_0)) + c \leq KS_o(A_0) + c.$$

Определение 3. Пусть существует алгоритм O , который максимально сжимает себя. Предположим, что принципиально новый алгоритм A имеет колмогоровскую сложность, меньшую колмогоровской сложности алгоритма O , порождающего A . Тогда для получения новых алгоритмов класса A необходимо уменьшить колмогоровскую сложность алгоритмов O . Более того, алгоритм O можно использовать для максимального сжатия самого себя только один раз. Для этого процесса требуется добавление внешней информации.

Итак, $O(O) = O'$ и $O'(O') = O''$, и $KS(O') < KS(O) + c$. Но поскольку $O'(O') = O''$, то $KS(O'(O')) = KS(O')$. Очевидно, что в будущем, чтобы получить $KS(O'')$ меньше $KS(O')$, необходима дополнительная модификация алгоритма O' с помощью новой информации.

В закрытой информационной системе «агент — окружающая среда» такая информация может быть стохастической составляющей, присутствующей либо в агенте, либо в среде.

Таким образом, интеллект требует максимизации информации об окружающей среде в содержащихся алгоритмах, максимального сжатия этих алгоритмов, а также наличия стохастической составляющей в системе «агент — окружающая среда».

3.2. Интеллектуальная система имеет сознание как интеграцию информации о себе, окружающей среде и положении агента в окружающей среде

Исходя из того факта, что сильный интеллект обладает способностью к адаптации, эмерджентностью и агентностью, необходимо, чтобы он обладал сознанием как инструментом внутренней самоорганизации [12]. Без этого необходимо было бы включить сущность Демона Максвелла, который выполнял бы эти функции. Это противоречит принципу бритвы Оккама, поскольку подразумевает включение ненужного объекта. Эта степень самоорганизации подразумевает высокий уровень взаимодействия сложных информационных процессов, относящихся к сильному интеллекту. Таким образом, можно предположить, что общая интеллектуальная система имеет сознание, а сознательная система имеет сильный интеллект.

Одной из популярных теорий, количественно определяющих сознание, является теория интегрированного сознания (ТИС) [13]. Ее авторы определяют сознание как единственное, что непосредственно и немедленно нам известно, и опосредует наше знание внешнего мира. Аксиомы ТИС предназначены для выявления существенных свойств опыта и включают: внутреннее существование, структуру, информацию, интеграцию и исключение. Итак, ТИС определяет, что сознание существует, структурировано, специфично, унифицировано и определено.

Таким образом, ТИС позволяет в некоторой степени количественно оценить степень системного сознания, но имеет следующие недостатки. Согласно ТИС сознание — это «вещь в себе», независимая от агентности и степени автономности агента. В этом подходе сознание возникает там, где имеется достаточная степень интеграции входящих информационных потоков. В то же время очевидно, что сознание неотделимо от агента и его интеллекта и основывается на автономности его восприятия.

В ТИС не представлен процесс самоорганизации сознания, сущность и необходимость его происхождения никак не объясняются. ТИС аксиоматически подтверждает уникальность индивидуального сознательного опыта, в то время как невозможно экспериментально оценить этот субъективный опыт и измерить степень различия в индивидуальном сознательном восприятии одних и тех же явлений разными наблюдателями.

Таким образом, основным недостатком ТИС в терминах сильного искусственного интеллекта является отсутствие эгоцентрического и четкого разделения между агентом и внешней средой. Чтобы преодолеть этот недостаток, авторы предлагают рассматривать сознание в первую очередь с позиции агентности, а также отношений между агентом и окружающей средой. В связи с этим авторы предлагают расширить понятие сознания из ТИС как **возможность интегрировать информацию агента о себе, окружающей среде и положении агента в этой среде** (рис. 2).

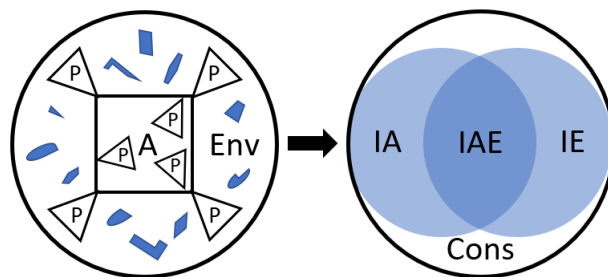


Рисунок 2. Сознание как интеграция информации агента о себе, окружающей среде и положении агента в этой среде. *A* — агент; *P* — поля восприятия агента (внутренние или внешние); *Env* — окружающая среда с некоторыми объектами; *IA* — информация об агенте; *IE* — информация об окружающей среде; *IAE* — информация о положении агента в окружающей среде; *Cons* — сознательный опыт

Агент, обладающий восприятием своего собственного состояния, в том числе умственного, находится в окружающей среде, взаимодействует с ней, а также, воспринимает состояние среды и результаты взаимодействия с ней. Таким образом, сознание представляется как направленный процесс обновления информации об окружающей среде и роли агента в ней.

3.3. Интеллект имеет аппарат для создания сложных абстрактных моделей

Для независимого развития и улучшения условий автономности агента сильный искусственный интеллект должен включать в себя аппарат для создания новых алгоритмов и сложных алгоритмов восприятия. Эти подсистемы можно обозначить как мышление, восприятие и воображение.

Мышление — это способность рассуждать, являющаяся процессом отражения объективной реальности в представлениях, суждениях, концепциях. Восприятие — это непосредственное чувственное отражение реальности в уме, способность воспринимать, различать и накапливать явления внешнего мира. Воображение — это свойство создавать образы, представления, идеи объектов в сознании, которые в опыте в целостном виде не воспринимались или не могут восприниматься посредством органов чувств. Воображение неразрывно связано с мышлением. По словам Выготского Л. С. [14], можно говорить о единстве этих двух процессов. Восприятие — это алгоритм для преобразования и распознавания входной информации.

На рис. 1 алгоритмы мышления и воображения находятся в блоке АГ (генератор алгоритмов). Восприятие находится в общем блоке алгоритмов. Эти подсистемы могут использоваться как для самооценки агента, так и для построения ментальных моделей (воображения) и оценки среды (мышления).

Эти три процесса вместе представляют собой аппарат для создания сложных абстрактных моделей. Можно воспринимать его как поэтапный процесс: во-первых, распознавание происходит с помощью алгоритмов распознавания, затем интегрируется блок мышления, который обрабатывает распознанную информацию и превращает ее в интеллектуальные конструкции, которые могут стать основой для воображения, необходимого для создания сложных абстрактных моделей.

3.4. Интеллект характеризуется эмерджентной самоорганизацией

Аппарат для создания сложных абстрактных моделей может функционировать автономно только из-за его эмерджентного характера. Таким образом, алгоритмы построения моделей также должны создавать себя сами. Это требует наличия множества отдельных частей системы «эгоистичного сильного искусственного интеллекта», которые будут работать вместе, а в процессе сотрудничества и конкуренции создадут новые алгоритмы.

Известно, что биологический мозг состоит из нейронов. Как показано в работе [15], один нейрон представляет собой систему с автономными гомеостатическими целями и осуществляет целенаправленное взаимодействие с окружающей средой. Примерами таких самоорганизующихся целенаправленных взаимодействий в нейронных системах являются байесовский вывод в нейронах неокортекса [16] и поддержание гомеостаза частоты спайкового разряда, включая изменение заданного значения гомеостаза частоты спайкового разряда в системе «нейрон-глия». В мозге также наблюдается самоорганизующаяся критичность [17], которая по определению требует, чтобы система имела много составляющих. Также была предложена модель самоорганизующегося управления, основанная на гомеостатических нейронных сетях, где каждый нейрон обладал определенной степенью активности в смысле необходимости сохранения собственного гомеостаза и принятия решений. Эта модель показала высокую степень адаптивности при решении задач управления агентами в стохастической среде по сравнению с обучением с подкреплением [15].

Модель такого рода может восприниматься как шаг к построению и реализации теории социальных нейронов. Авторы предлагают теорию, в которой нейроны взаимодействуют друг с другом как социальные агенты, осуществляющие сотрудничество и конкуренцию. Такой подход согласуется с работой [18].

4. Заключение

В данной статье предложен новый подход к интеллекту человеческого уровня под названием «эгоистичный сильный искусственный интеллект». Он фокусируется на

автономии и агентности интеллектуальной системы и ее функционировании, исходя из собственных интересов. Предложен полный минимальный список свойств, присущих интеллектуальным системам человеческого уровня. Интеллект рассматривается с алгоритмической точки зрения и с позиции создания принципиально новых алгоритмов сжатия. Изучено сознание как важное свойство интеллектуальной системы. Предложен трехкомпонентный подход к количественной оценке сознания. Показано, что интеллектуальная система должна состоять из отдельных самоорганизующихся частей. Предложены четыре постулата интеллекта. Предполагается, что предложенный в работе подход позволит уточнить и прояснить путь к созданию сильного искусственного интеллекта.

Литература

- [1] *Linde A.* Sinks in the landscape, boltzmann brains, and the cosmological constant problem // *Journal of Cosmology and Astroparticle Physics*. 2007. Vol. 2007. No. 01. P. 022.
- [2] *Kohler W.* Gestalt psychology. — NY : Liveright Co, 1947.
- [3] *Wertheimer W.* Productive thinking. — NY : Harper, 1945.
- [4] *Glaser R.* Education and thinking: The role of knowledge // *Amer Psychologist*. 1984. Vol. 39. No. 2. P. 93–104.
- [5] *Piaget J.* Development and learning // *Piaget Rediscovered: A Report on the Conference of Cognitive Studies and Curriculum Development / Eds.: Ripple R. E., Rockcastle V. N.* — NY : Cornell University, Ithaca, 1964. P. 7–20.
- [6] *Gottfredson L. S.* Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography // *Intelligence*. 1997. Vol. 24. No. 1. P. 13–23.
- [7] *Goertzel B., Pennachin C., Geisweiller N.* Engineering General Intelligence, Part 2 The Cog-Prime Architecture for Integrative, Embodied AGI. — NY : Springer USA, 2014.
- [8] *Weinbaum D. W., Veitas V.* Open ended intelligence // *Journal of Experimental Theoretical Artificial Intelligence*. 2017. Vol. 29. No. 2. P. 371–396.
- [9] *Zhdanov A. A.* About an autonomous adaptive control methodology // *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC)*. Gaithersburg, MD, 1998. P. 227–232.
- [10] *Dowe D. L., Hernandez-Orallo J., Das P. K.* Compression and intelligence: Social environments and communication // *Artificial General Intelligence. AGI 2011*. Vol. 6830 of LNCS. — Springer Berlin Heidelberg, 2011. P. 204–211.
- [11] *Shen A., Uspensky V. A., Vereshchangin N.* Kolmogorov Complexity and Algorithmic Randomness. American Mathematical Society // *Mathematical Surveys and Monographs*. 2017. Vol. 220.
- [12] *Peters F.* Consciousness as recursive, spatiotemporal self-location // *Psychological research*. 2009. Vol. 74. P. 407–421.

- [13] *Tononi G.* Consciousness as an integral information: a provisional manifesto // *The Biological Bulletin*. 2008. Vol. 215. No. 3. P. 216–242.
- [14] *Выготский Л. С.* Воображение и творчество в детском возрасте: психологический очерк: книга для учителя. — 3-е изд. — М. : Просвещение, 1991.
- [15] *Nikitin O., Lukyanova O.* Control of an agent in the multi-goal environment with homeostasis-based neural network // *Procedia Computer Science*. 2018. Vol. 123. P. 321–327.
- [16] *Deneve S.* Bayesian inference in spiking neurons // *Advances in Neural Information Processing Systems*. Vol. 17. — MIT Press, 2005. P. 353–360.
- [17] *Hesse J., Gross T.* Self-organized criticality as a fundamental property of neural systems // *Frontiers in Systems Neuroscience*. 2014. Vol. 8. P. 166.
- [18] *Seung H. S.* Learning in spiking neural networks by reinforcement of stochastic synaptic transmission // *Neuron*. 2003. Vol. 40. No. 6. P. 1063–1073.

Авторы:

Ольга Александровна Лукьянова — научный сотрудник, лаборатория «Информационно-телекоммуникационных систем», Вычислительный Центр ДВО РАН

Олег Юрьевич Никитин — научный сотрудник, лаборатория «Информационно-телекоммуникационных систем», Вычислительный Центр ДВО РАН

Selfish General Intelligence

O. A. Lukyanova, O. Y. Nikitin

Computing Center, FEB RAS, 65, Kim Yu Chen Street, Khabarovsk, Russia, 680000
e-mail: ollukyan@gmail.com, olegioner@gmail.com

Abstract. In this paper, we attempted to formulate a complete minimal list of properties corresponding to human-level intelligent systems. Intelligence was presented from the algorithmic point of view and considered as the ability to generate fundamentally new compression algorithms. Consciousness was studied as an important property of an intelligent system. The approach to consciousness as a mechanism for integrating information from three components was proposed. Also, emergent generation of fundamentally new algorithms was suggested. It was shown that the intelligent system should consist of separate self-organizing parts. This approach to the evaluation and development of human-level intelligence is called Selfish General Intelligence (SGI). The four proposed postulates of intelligence could help to clarify the understanding of AGI.

Keywords: selfish general intelligence, artificial general intelligence, Kolmogorov complexity, consciousness, self-organization, compression algorithm, autonomy, agent.

References

- [1] *Linde A.* (2007) *Journal of Cosmology and Astroparticle Physics*, **2007**(01):022.
- [2] *Kohler W.* (1947) *Gestalt psychology*. New York: Liveright Co.
- [3] *Wertheimer W.* (1945) *Productive thinking*. New York: Harper.
- [4] *Glaser R.* (1984) *Amer Psychologist*, **39**(2):93–104.

- [5] Piaget J. (1964) Development and learning. In *Piaget Rediscovered: A Report on the Conference of Cognitive Studies and Curriculum Development*. Ripple R. E., Rockcastle V. N., editors. P. 7–20. Cornell University, Ithaca, NY.
- [6] Gottfredson L. S. (1997) *Intelligence*, **24**(1):13–23.
- [7] Goertzel B., Pennachin C., Geisweiller N. (2014) *Engineering General Intelligence*, Part 2 The Cog-Prime Architecture for Integrative, Embodied AGI. New York, Springer USA.
- [8] Weinbaum D. W., Veitas V. (2017) *J. of Experimental Theoretical Artificial Intelligence*, **29**(2):371–396.
- [9] Zhdanov A. A. (1998) About an autonomous adaptive control methodology. In *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC)*, Gaithersburg, MD. P 227–232.
- [10] Dowe D. L., Hernandez-Orallo J., Das P. K. (2011) Compression and intelligence: Social environments and communication. In *Artificial General Intelligence*, Schmidhuber J., Thirsson K. R., Looks, M. (eds.) AGI 2011. Vol. 6830 of LNCS. P. 204–211. Springer Berlin Heidelberg.
- [11] Shen A., Uspensky V. A., Vereshchangin N. (2017) *Kolmogorov Complexity and Algorithmic Randomness*. American Mathematical Society, Mathematical Surveys and Monographs 220.
- [12] Peters F. (2009) *Psychological research*, **74**:407–421.
- [13] Tononi G. (2008) *The Biological Bulletin*, **215**(3):216–242.
- [14] Vygotsky L. S. (1991) *Imagination and creativity in childhood*. Education, Moscow. [In Rus]
- [15] Nikitin O., Lukyanova O. (2018) *Procedia Computer Science*, **123**:321–327.
- [16] Deneve S. (2005) Bayesian inference in spiking neurons. In *Advances in Neural Information Processing Systems*, Saul L. K., Weiss Y., Bottou L., (eds.), 17. P. 353–360. MIT Press.
- [17] Hesse J., Gross T. (2014) *Frontiers in Systems Neuroscience*, **8**:166.
- [18] Seung H. S. (2003) *Neuron*, **40**(6):1063–1073.