



Artificial societies. 2013-2024

ISSN 2077-5180

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

All right reserved

Issue 3 Volume 13. 2018

Autonomous Agent of an Artificial Society as “Whole”, “Alive”, “Conscious”: Attempt of Modeling

Evgeny Grishin

Independent researcher and inventor

Russian Federation, Moscow

Abstract

An approach is proposed meaningfully interpreting the concept of an autonomous agent - an element of the model of an “artificial society” as something “whole”, “alive” and “conscious”. The goal is to lay the foundation for computer modeling of the phenomenon of “life and conscious activity” of an agent in “society as a multi-agent system”. For analysis the concepts “universal Turing machine (UMT)” and the principle of UMT self-applicability as a way of its self-programming were used. It is shown that as a result, the problem of algorithmic non-computability of the UMT self-management problem can be removed.

Keywords list (en): artificial life, machine consciousness, Turing machine, behavior

Date of publication: 03.07.2018

Citation link:

Grishin E. Autonomous Agent of an Artificial Society as “Whole”, “Alive”, “Conscious”: Attempt of Modeling // Artificial societies. – 2018. – V. 13. – Issue 3. URL: <https://artsoc.jes.su/s207751800000128-9-1/>. DOI: 10.18254/S0000128-9-1

1 Попытаемся ответить на концептуальный вопрос: какова должна быть общая структура и функции автономного агента как элемента некоего «искусственного общества»? Достаточно ли представлять агента как узел в некоей сети, реагирующий тем или иным образом на внешние стимулы, и не должен ли агент представлять собой нечто большее - «целое», «живое» и «сознающее»?

2 Технологически легко обеспечить простую репликацию (порождение) одним автоматом других автоматов [1]. При этом можно организовать «эволюцию» популяции автоматов [2]. Существует множество примеров, представляющих с разной степенью адекватности подобные модели как модели «искусственной жизни». Но понимать этот термин буквально как искусственное воспроизведение жизни, наверное, было бы натяжкой.

3 В чистом виде это не более чем процессы генерации автоматов некоей операционной системой (или многократное копирование автоматов). При этом каждый из автоматов сам по себе не является (не обязан быть) чем-то живым, и не обязан обладать самостоятельной функцией самокопирования. Достаточно задать строгие или вероятностные правила «стимул-реакция» для моделирования взаимоотношений агентов [3].

4 Есть множество примеров процессов самоорганизации в *неживой* материи, успешно исследуемые *синергетикой* [2]. В рамках синергетики ищутся механизмы решения проблемы *целого*. Но есть ощущение, что это будет решение проблемы *целого* для *неживого*, так как не определена разница между живым и неживым.

5 Для решения проблемы *целого* для *живого* (биологического и социального) может помочь применение принципа *заимствования*, сформулированного В.А. Лефевром [4]. С одной стороны, этот принцип предполагает существование таких объектов, в *структуре* которых имеется *проект строительства самого себя*. С другой стороны, предполагается наличие *механизма* переноса (заимствования) проекта при порождении одним объектом другого, подобного первому. Принцип *заимствования*, на наш взгляд, может быть базовым признаком отличия «квазживых субъектов» от «неживых» объектов, по крайней мере, при компьютерном моделировании жизни.

6 Представляется, что для адекватного моделирования жизни (т.е., практически, для создания популяции неких искусственных организмов), необходимо, чтобы сам базовый элемент порождаемого множества уже был «живым», или, по крайней мере, «квазживым», в рамках принципа *заимствования*. Попытаемся (самостоятельно от автора [4]) развернуть собственное понимание принципа заимствования, сохраняя в неприкосновенности исходную посылку. Напомним, что данный принцип, в общем, предполагает наличие у субъекта проекта самого себя, т.е. образца и механизма строительства самого себя «завтрашнего», или порождения другого субъекта, подобного первому, в соответствии с этим образцом.

7 Перечислим в порядке убывания общности перечень качеств, интуитивно нами приписываемый «живому» и «сознающему» субъекту, образец которого следует использовать для строительства себя «завтрашнего», или другого подобного субъекта:

- 8 • Свойство *репликации*, т.е. самостоятельная способность к *порождению* одним субъектом другого, себе подобного, или самого себя «завтрашнего».
- «*Физиологическая*» *активность*. Имеется в виду способность субъекта к *накоплению* из среды и *циклическому* преобразованию энергии как расходу

одного вида материального ресурса и выработке другого вида материального ресурса. Иначе говоря, это - требование *метаболизма*, или *обмена веществ*.

- *Потребность* как способ проявления «физиологической» активности в виде стремления к пополнению какого-либо жизненного материального ресурса с наибольшей текущей нехваткой относительно некоей нормы, аperiodически переключаемого на другой ресурс с наибольшей нехваткой.
- *Деятельность* как некая внешняя *активность субъекта в среде*, результат которой есть последовательное удовлетворение его неудовлетворённых *Потребностей*. Деятельность есть триада: *Осознание*, *Поведение* и *Супервизор-Переключатель*. *Осознание* есть *Анализ* и *Синтез*, которые последовательно исполняются соответствующими рекурсивными алгоритмами Супервизора: *a) Анализ* расхождения *прошлого* Плана и Факта, построение Тенденций и Прогноз *будущих* последствий, Формулирование Проблемы, Постановка Задачи (Цели); *b) Синтез* *будущего* Плана устранения расхождения План-Факт. В то же время Осознание осуществляется на фоне исполнения пошагового рекурсивного алгоритма *текущего Поведения* под контролем Поискового Плана Супервизора. Разработанный в Осознании *будущий* План устранения расхождения План-Факт *переключается* в *текущее Поведение* и также ставится на контроль исполнения Супервизором. Неисполненный План в *текущем Поведении переключается Супервизором в Осознание* и становится *Прошлым*. Обобщённо говоря, Супервизором по прерыванию осуществляется *одновременное круговое переключение* всех трёх Планов: Текущий План становится *Прошлым* и анализируется, из проанализированного позапрошлого Плана синтезируется План Будущего, предыдущий План Будущего становится Текущим Планом.
- *Проблема* как оценка вероятных потерь ресурсов, которые могут последовать из-за помех в осуществлении любой *Деятельности*, оценка готовности к затратам ресурсов на устранение помех, и распределение затрат на каждом из вышеперечисленных этапов. Неготовность к затратам ресурсов на устранение помех есть отсутствие Проблемы.
- *План поведения* (список ролей и последовательность действий каждого исполнителя и *График* их взаимодействия) как путь устранения *Проблемы*. Сам факт появления плана у субъекта уже есть способ психологического разрешения проблемы.
- *Принятие решения* как процесс генерации вариантов или поиска источников генерации, и выбора вариантов Исполнителей Ролей с учётом вероятностей их исполнительности и затратности, на каждом этапе *Графика*.
- *Информация* как *Ответ* на *Вопрос*, заданный *Субъектом*, в виде: Кто/Что, Где, Когда, Почему, Зачем, Сколько, Кто за Что и Как отвечает. Вопрос появляется тогда, когда у Субъекта есть необходимость принять какое-то *Решение* и нет источника вариантов или требуемых знаний для оценки появляющихся вариантов. Ответ может быть получен им или от другого Субъекта, или от себя самого на основе собственного опыта. *Без вопроса Субъекта Информации не существует!*

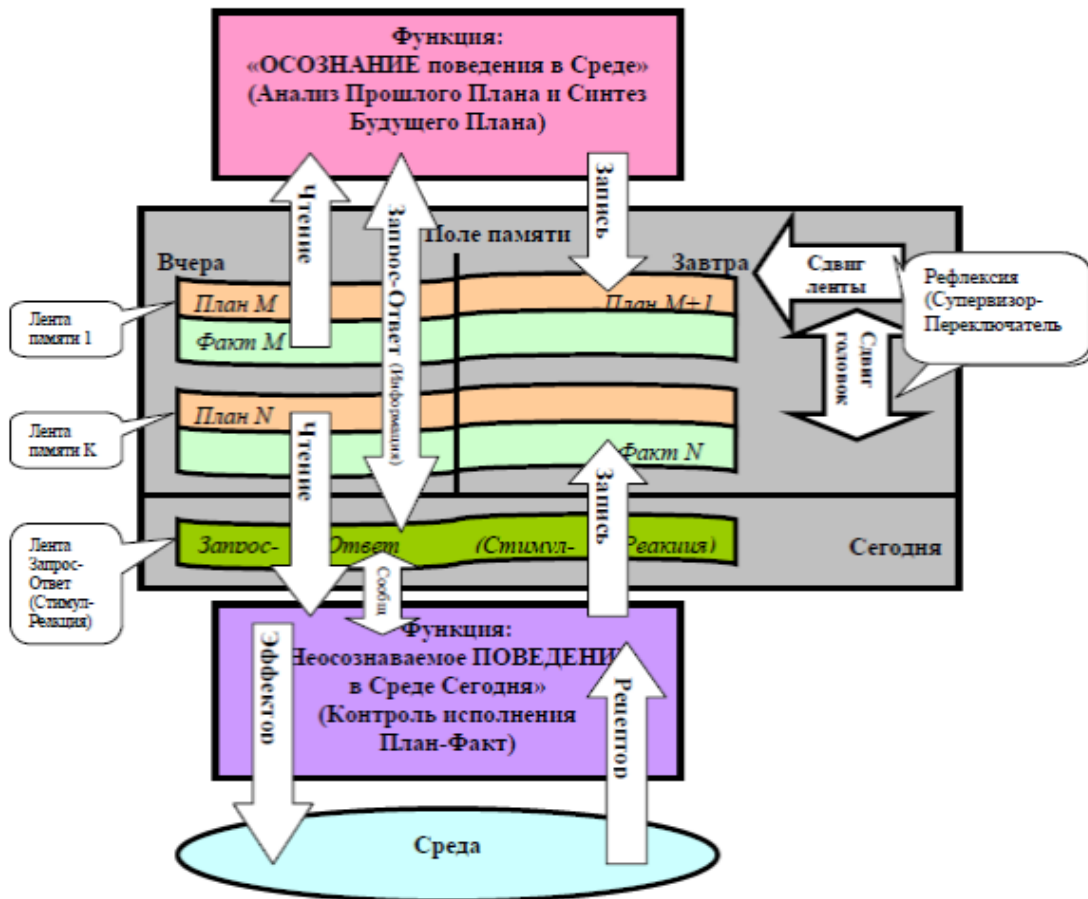


Рис. 1. Рефлексивная машина.

11 В свою очередь, Проектный механизм есть набор инструментов для строительства по данному образцу, который, по нашему мнению, адекватно может быть представлен *рефлексивной машиной* (рис. 1). В целом, рефлексивная машина есть система из двух универсальных машин Тьюринга, параллельно работающих на общем поле памяти. Термин «рефлексивная машина» впервые задан автором в [6] как инвариантная модель *технологии информационной обратной связи* в деятельности агента, обратной связи между осознанием и поведением агента. Более основательно структура и функционирование рефлексивной машины рассмотрено в специальной работе [7].

12 В рамках исследовательского направления «искусственный интеллект» известны попытки анализа проблемы «самоприменимости» универсальной машины Тьюринга (УМТ) [5].

13 Вкратце, самоприменимость представляет собой такой режим работы УМТ, когда она генерирует на ленте памяти не просто данные, а программный код собственного устройства управления, который сама же далее читает и интерпретирует (исполняет). При этом важно, что самоприменимость рассматривается как алгоритмический вариант реализации функции «машинного сознания». Цель такого анализа – найти подтверждение или опровержение тезиса об алгоритмической невычислимости самой задачи самоприменимости УМТ, а значит, и подтверждение или опровержение по данной причине тезиса о невозможности машинного «интеллекта», машинного «сознания».

14 На наш взгляд, самоприменимость УМТ, рассмотренная вышеизложенным способом, только и может, что привести к выводу об алгоритмической невычислимости задачи, а значит, и к нерешаемости проблемы «машинного сознания» таким способом. Причина тому – тривиальное заикливание ввиду самовывоза начальной инструкции управляющей программы.

15 Предложим иной подход, в результате которого может быть снята проблема алгоритмической невычислимости задачи самоприменимости УМТ. Иначе говоря, покажем, что существуют условия, при которых допустима самоприменимость УМТ, а значит, и теоретически не запрещена алгоритмическая реализация «машинного сознания» на этом принципе.

16 Предположим, что некая УМТ в течение какого-то периода времени своей работы генерирует *не* собственный программный код, а некие данные, которые будет читать в следующем периоде *другая* УМТ, работающая параллельно с первой на этом же поле памяти. И восприниматься эти данные *другой* УМТ будут не как инструкции программы, а как некий *план*, в соответствии с которым её управляющая программа должна будет включать и выключать требуемые исполнительные алгоритмы своего поведения (рецепторы и эффекторы) в требуемое время.

17 Далее, эта *вторая* УМТ, читая и исполняя в следующем периоде полученный *план*, запишет в общее поле памяти некий *факт* выполнения этого плана. Этот фактический результат выполнения плана *второй* УМТ в очередном периоде прочтёт *первая* УМТ, и если план не выполнен, включит соответствующий алгоритм генерации нового плана. Таким образом, выше описана *рекурсивная* процедура: генерация данных одной УМТ и чтение этих данных другой УМТ.

18 С целью более эффективного донесения сути подхода дальнейшее изложение части материала осуществим в виде диалога «вопрос-ответ» автора с самим собой.

19 **Вопрос:** а что делает другая УМТ в тот же период, когда одна УМТ читает и генерирует данные? **Ответ:** делает то же самое - читает и генерирует данные!

20 **Вопрос:** какие данные читает и генерирует каждая из УМТ в текущий период времени? **Ответ:** в текущем периоде первая УМТ *читает* данные, сгенерированные и записанные второй УМТ в *предыдущем* периоде. Вторая УМТ *читает* данные, сгенерированные и записанные первой УМТ тоже в *предыдущем* периоде! А также - в текущем периоде первая УМТ *генерирует* и записывает данные, которые будут прочтены второй УМТ в *будущем* периоде. Вторая УМТ *генерирует* и записывает данные, которые будут прочтены первой УМТ в *будущем* периоде.

21 **Вопрос:** в *текущем периоде* обмен данными между двумя УМТ через общее поле памяти есть? **Ответ:** да, есть! Этот обмен представляет собой следующее: первая УМТ (Осознание поведения в Среде) формулирует на специальной ленте поля памяти *вопросы*. Одновременно вторая УМТ (Поведение в Среде), действуя по своему поисковому плану и алгоритму, включает эффекторы и

рецепторы связи со Средой, получает сообщения от Среды и записывает их на эту же специальную ленту поля памяти.

22 Если пришедшие данные есть *ответы* на вопросы (*информация*), то первая УМТ (Осознание) читает и воспринимает её как характеристику вариантов для принятия решений (в рамках осознания причин срыва предыдущего плана, или в рамках разрабатываемого ею плана будущего поведения).

23 Одновременно вторая УМТ (Поведение в Среде) использует эти же сообщения Среды, полученные от рецепторов, как *стимулы* для непосредственного включения соответствующих *реакций* эффекторов в Среде в соответствии с исполняемым алгоритмом поведения.

24 И вопросы, и ответы записываются на специальной ленте памяти в «текущий регистр» прямого доступа и в текущем периоде доступны для чтения непосредственно. В следующем же периоде «текущий регистр» будет освобождён, данные будут сохранены на специальной ленте памяти (Бессознательное), и технология доступа к ним усложнится. В «текущий регистр» будут записываться и читаться новые вопросы и ответы (стимулы и реакции) нового текущего периода.

25 **Вопрос:** Что же конкретно представляет собой *предмет осознания* в проявлении системы вообще? **Ответ:** В данном исследовании определяется, что предметом «осознания» системы является только *Среда и собственное поведение системы, отраженное в Среде, и только в прошедший период*, и воспринимаемые только как соотнесение плановых и фактических состояний, записанных на *временной* ленте памяти как прошлое и записываемых на *временную* ленту памяти как план и алгоритм на будущее!

26 Собственно внутренние алгоритмы системы (чтение, запись данных, рекурсивные алгоритмы анализа, синтеза и контроля планов, обмен вопрос-ответ, рефлексивные прерывания и переключения областей памяти), *принципиально не становятся* предметом осознания системой, и являются результатом осознания и модификации проекта в целом, что есть прерогатива *конструктора!*

27 Среда и поведение системы *в текущий период* существования отражаются как факт на участке поля памяти, называемом «будущее». Поэтому *в текущем периоде* самой системой они также *принципиально неосознаваемы*, и смогут быть прочитаны только в следующих периодах осознания, когда участок памяти «будущее» перейдёт в категорию «прошлое». Оно (поведение) всегда осознаётся, лишь становясь прошлым, а текущее поведение будет новым предметом осознания «постфактум», лишь в предстоящем периоде.

28 **Вопрос:** Что же представляет собой *средство для осознания* системой её собственного поведения? **Ответ:** В данном исследовании определяется, что *средством для осознания* (осуществляемого всегда и только в текущем периоде) системой её собственного поведения, является *информация*, т.е. только такие сообщения из Среды, которые расцениваются как *ответ на вопрос*:

- 29
- о причинах невыполнения «план-факт» поведения в *прошедший период*,
 - о вариантах поведения для принятия решений о плане в *будущем периоде*.

30 **Вопрос:** Что представляет собой *осознание как процесс*? **Ответ:** В соответствии с вводимым нами узким определением, осознание как процесс имеет две составляющие, внешнюю и внутреннюю:

- 31 • внешняя составляющая - *поведенческий поиск и отбор персонажем сообщений Среды как ответов на вопросы*, которые формулируются им в текущем периоде. Ответы также получаются в текущем периоде, и должны быть классифицированы по релевантности, оперативности, актуальности и достоверности (поток сообщений характеризуется большим «шумом» и большим разнообразием характеристик источников - временная задержка, достоверность и давность).
- внутренняя составляющая - *восприятие персонажем череды получаемых сообщений Среды как альтернативных вариантов для принятия решений*: - при выяснении *причин невыполнения «план-факт»* конкретного поведения системы в *прошедший период*; - с целью разработки *плана будущего поведения* системы, устраняющего причину невыполнения «план-факт» прошлого её поведения.

32 Именно в этом, и пока ни в каком другом, смысле нами рассматривается *Осознание* персонажем Среды и его собственного поведения в Средe, происходящее в текущем периоде!

33 Вспомним, что те же самые сообщения рецепторов *воспринимаются* УМТ (Поведение) как *стимулы* для запуска *реакций* эффекторов, функционирующих в Средe! Однако этот процесс нами относится не к *феномену Осознания*, а к *феномену неосознаваемого в текущем периоде Поведения*. Сочетание же Осознания и Поведения, по мнению автора, и образуют феномен Целого (в соответствии с «принципом дополнительности» Н.Бора)!

34 Таким образом, общее поле памяти представляет собой *структуру*, включающую следующее:

- 35 • Набор лент *временной памяти* (Осознаваемое); каждая из них разделена границей на отрезки «прошлое» (для чтения данных «план-факт») и «будущее» (для записи данных «план-факт»). Данные представляют собой имена, наименования, обозначения, даты и числа, которые связаны с создаваемыми рецепторами образами объектов Среды.
- *Текущий регистр и вневременная лента памяти* (Бессознательное) для чтения и записи данных (образов объектов Среды), которые воспринимаются как «запрос-ответ» при текущем принятии решений в новых планах, и как «стимул-реакция» при текущем обеспечении условных переходов в алгоритмах поведения эффекторов и рецепторов.

36 Рефлексия есть... прерывание Супервизором функций Поведения и Осознания, временной сдвиг Лент Памяти и Головок чтения/записи, смена Планов и Алгоритмов Поведения и Осознания, а также очередное включение функций Поведения и Осознания!

37 Поведение системы из двух УМТ и общего поля памяти заключается в следующем. В конце каждого временного периода осуществляется *горизонтальный* СДВИГ ВЛЕВО на один сектор двух временных лент, над которыми стоят пишущие и читающие головки обеих УМТ. Если при этом УМТ-Поведение, читая данные на ленте, констатирует неисполнения плана деятельности в прошедшем периоде, то она осуществляет *вертикальный* СДВИГ своих пишущих и читающих головок на другую ленту с готовым новым планом. В свою очередь, УМТ-Осознание осуществляет *вертикальный* СДВИГ своих пишущих и читающих головок на ленту с неисполненным планом. Если УМТ-Поведение констатирует исполнение плана в прошедшем периоде, *вертикальных* сдвигов не происходит, и процесс исполнения плана продолжается.

38 Описанный процесс представляет собой, по сути, «машинную технологию» рефлексии, реализуемую в системе двух УМТ, которая есть:

- 39
- периодический сдвиг влево временной ленты памяти конкретной деятельности под головками УМТ-Поведение,
 - констатация УМТ-Поведение неисполнения плана деятельности,
 - вертикальный сдвиг самих головок УМТ-Поведение на ленту памяти другой (поисковой) деятельности с готовым планом,
 - одновременный вертикальный сдвиг головок УМТ-Осознание на ленту памяти с неисполненным планом.

40 Продолжим собственную расшифровку «принципа заимствования» [4] с учётом предыдущих соображений о связке двух УМТ. В соответствии с предыдущим, термин «субъект содержит проект самого себя» необходимо должен включать два аспекта рассмотрения субъектом самого себя:

- 41
- Текущее *осознание* субъектом самого себя как результат *прошлой* реализации некоего проекта себя, разработанного и реализованного в *прошедшем* периоде времени. Осознание есть констатация факта исполнения или неисполнения плана реализации прошлого проекта и формулирование в связи с этим проблемы, которую нужно решать.
 - Текущее *представление* субъекта о себе («образец самого себя») в *будущем* периоде времени, завершаемое разработкой плана устранения осознанной проблемы.
 - Проявление же субъекта в текущий период времени есть ни что иное, как *строительство себя по ранее выработанному образцу* (т.е., текущее поведение по плану). И это проявление есть принципиально *неосознаваемое поведение* в текущий период времени, обеспечивающее информацией функцию Осознания.

42 Теперь совместим понятия:

- 43
- Персонаж – это пара УМТ с общим полем памяти.
 - УМТ-1 – это функция «Осознание».
 - УМТ-2 – это функция «Поведение».
 - «Проект самого себя» - это План М будущего поведения персонажа.

- «Реализация проекта самого себя» - это Факт прошлого исполнения Плана М-1 персонажа. План и Факт, расположенные на одной ленте памяти, имеют отношение к одной конкретной деятельности персонажа.

44 Функция «Осознание» имеет в качестве области определения «прошлое» поля памяти, где она читает данные Факта исполнения Плана М-1. В качестве области значений она имеет «будущее» поля памяти, где она записывает как данные План М (Табл. 1). Функция «Поведение» имеет в качестве области определения «прошлое» поля памяти, где она читает данные Плана N-1. В качестве области значений она имеет «будущее» поля памяти, где она записывает как данные Факт исполнения Плана N-1».

45 Таблица 1

Функция	Источник и хранилище данных	Прошлое (область определения)		Будущее (область значений)			
		Чтение	Запись	Период 1	Период 2	Период 3	Период 4
Функция «Осознание»	Лента памяти 1	План Факт	План М-1	Факт	План М		
Функция «Поведение»	Лента памяти К	План Факт		План N-1	Факт N-1		

46 Функция «Осознание» Лента «Бессознательное» Функция «Поведение»

Регистр Текущего Периода i

Формулирование и запоминание *Запроса* к Среде. Чтение и интерпретация *Ответа* из Среды (*Информация*)

Восприятие *Стимула* из Среды. Формирование *Реакции* в Среде.

47 Из вышеизложенного получаются следующие выводы:

- 48 • Для решения проблемы моделирования *целого* для *живого* (биологического и социального) может помочь применение принципа *заимствования*, сформулированного В.А. Лефевром [4]. С одной стороны, этот принцип предполагает существование таких объектов, в *структуре* которых имеется *проект строительства самого себя*. С другой стороны, предполагается наличие *механизма* переноса (заимствования) проекта при порождении одним объектом другого, подобного первому. Автором предложен вариант реализации *Проектного образца* и *Проектного механизма* в виде *Рефлексивной машины*.
- Самоприменимость УМТ как принципиальная модель «самоуправления» УМТ сохраняется, но «самопрограммирование» одной УМТ заменяется на «самопланирование» системы двух УМТ.
- Замена самопрограммирования на самопланирование означает принципиальную ликвидацию «зацикливания» УМТ. При самопрограммировании «зацикливание» происходит ввиду того, что передача управления осуществляется рекурсивно (т.е., не по внешнему прерыванию, а по внутренней логике одной и той же программы). При «самопланировании» зацикливание не происходит ввиду того, что передача управления от одной УМТ другой осуществляется принудительно, по внешнему прерыванию от другой УМТ или от операционной системы. Следовательно, в такой

постановке снимается вопрос об алгоритмической невычислимости задачи самоприменимости УМТ.

- Значит, теоретически не закрывается дорога для реализации «машинного сознания» на принципе самоприменимости УМТ, правда, уже в рамках самопланирования системы двух УМТ на общем поле памяти.

⁴⁹ Автор данного текста вполне отдаёт себе отчёт в том, сколь непроста задача - донести своё представление о «машинном мышлении» до читателя. Слабым оправданием может быть лишь то, что автор вынужден применять мысль как инструмент для описания самого же используемого инструмента. А такой способ, как показал Гёдель, изначально чреват неполнотой. К сожалению, автору другие рациональные способы донесения мысли о мысли неизвестны. Правда, практика разработки автором компьютерного персонажа на вышеизложенных принципах показала высокую степень адекватности допущения о «квазиживости» и «квазиразумности» поведения и взаимодействия подобного автомата среди партнёров-людей.

Автономный агент искусственного общества как «целое», «живое», «сознающее»: попытка моделирования.

Гришин Евгений Александрович

Независимый исследователь и изобретатель

Российская Федерация, Москва

Аннотация

Предложен подход, содержательно интерпретирующий понятие автономного агента – элемента модели «искусственного общества» как нечто «целое», «живое» и «сознающее». Цель - заложить основу для компьютерного моделирования феномена «жизненной и сознательной активности» агента в «обществе как многоагентной системе». В качестве исходного материала для анализа привлекаются понятия «универсальная машина Тьюринга (УМТ)», а также принцип самоприменимости УМТ как способ её самопрограммирования. В качестве инструмента синтеза автором используются «принцип заимствования», предложенный В.А. Лефевром, а также предлагаемая автором модель рефлексивной машины и принцип самопланирования (взамен принципу самопрограммирования УМТ). Показано, что в результате может быть снята проблема алгоритмической невычислимости задачи самоуправления УМТ как помеха на пути алгоритмического варианта реализации функции «машинного сознания».

Ключевые слова: искусственная жизнь, машинное сознание, машина Тьюринга, поведение, осознание

Дата публикации: 03.07.2018

Ссылка для цитирования:

Гришин Е. А. Автономный агент искусственного общества как «целое», «живое», «сознающее»: попытка моделирования. // Искусственные общества. – 2018. – Т. 13. – Выпуск 3. URL: <https://artsoc.jes.su/s207751800000128-9-1/>. DOI: 10.18254/S0000128-9-1