

# Универсальный разум AGI For All (AGIFA)

Артюхов Виктор (@LiveBaster)

<https://github.com/LiveBaster/agifa>

Ссылка на первоначальную презентацию проекта AGIFA:

[https://github.com/LiveBaster/agifa/blob/main/docs/agi\\_for\\_all.pdf](https://github.com/LiveBaster/agifa/blob/main/docs/agi_for_all.pdf)

Цель проекта - разработать универсальный интеллектуальный движок, способный к самообучению, развитию и решению задач разного типа.

Движок один - задачи разные.

Далее, я буду рассказывать **о текущих достигнутых результатах проекта AGIFA** и путях дальнейшего его развития.

# Определение универсального разума (AGI)

По «объединённому и расширенному определению» от Ben Goertzel, Pei Wang, Shane Legg, Marcus Hutter — способность достигать сложных целей в различных сложных средах, в условиях ограниченных ресурсов, минимизируя риски, в том числе, обучаясь поведению в новых средах как с учителем, так и без него.

<https://golos.id/ru--tekhnologii/@aigents/opredeleniya-ii?invite=aigents>

Колонин Антон

# Что требуется сделать, по определению?

- способность достигать сложных целей в сложных средах
- возможность минимизировать риски и расход ресурсов
- обучение с учителем или без него

Самое непонятное здесь - что понимать под “сложностью”?

Интуитивно понятно, что если мало компонентов, из которых создаётся универсальный разум, то это “простая система”. Например, управлять одним квадрокоптером - это достаточно просто.

Если компонентов много, то это “сложная система”. Например, управлять 10000 квадрокоптеров - это много сложнее, чем одним.

Гипотеза: сложная система сможет достигать сложных целей в сложных средах.

Каким требованиям должны удовлетворять компоненты, чтобы из них была возможность построить сложную систему?

# Требования к компонентам универсального разума

1. **взаимосодействие** параллельных компонентов для достижения общего требуемого результата
2. **выбор** требуемого результата и **поиск** способа его достижения
3. **саморегуляция**
4. **подражание**

Гипотеза: Без реализации этих 4-х свойств, создать универсальный разум не получится.

Каждый Базовый Компонент Разума (БКР) должен обладать всеми 4-мя свойствами.

# Соответствие требований БКР целям из определения

1. **взаимосодействие** параллельных компонентов для достижения общего требуемого результата - требуется для создания сложной системы
2. **выбор** требуемого результата и **поиск** способа его достижения - требуется для достижения целей
3. **саморегуляция** - требуется для минимизации рисков и расхода ресурсов
4. **подражание** - требуется для обучения с учителем или без него

*Далее, я буду детализировать каждый из четырёх пунктов.*

# Взаимосодействие для получения общего результата

"На меня неизгладимое впечатление произвело одно случайное наблюдение в поле у пасущегося стада коров. Был жаркий летний день, и я заметил, как животные хлестали себя непрерывными ударами кончика хвоста по местам, на которые садились кровососущие насекомые – оводы. Но вот один овод, очевидно, сел и кусал на таком месте кожи, которое не могло быть, так сказать, "прострелено" ударом хвоста. Все туловище животного изогнулось дугой, голова сделала максимальный поворот в области шейных суставов, вытянулась шея, и язык, вытянувшись в неожиданно для меня длинную "палку", пытался достать то место туловища, где уселся овод. Все мышцы туловища находились в судорожном движении, однако все до единого сокращения были направлены в одну сторону – обеспечить прикосновение кончика языка к месту укуса. Одновременно подкожная мышца того места, на котором сидел овод, сильно сокращаясь, приближала овода к кончику языка. Здесь, таким образом, мы наблюдаем поразительное взаимосодействие самых разнообразных мышц тела для обеспечения сбрасывания овода с кожи. Вот такая обширная организация, включающая мышцы, нервы, рецепторы, нервные центры и т.д., и может быть названа системой, поскольку она приводит к получению конечного полезного результата."

"Анохин П.К. А69 Избранные труды: Кибернетика функциональных систем/Под ред. К.В. Судакова"

# Взаимосодействие

“Под функциональной системой мы понимаем такое сочетание процессов и механизмов, которое, *формируясь динамически* в зависимости от данной ситуации, непременно приводит к конечному приспособительному эффекту, полезному для организма как раз именно в этой ситуации. Из приведенной формулировки следует, что функциональная система может быть составлена из таких аппаратов и механизмов, которые могут быть весьма отдаленными в анатомическом отношении. Это значит, что состав функциональной системы и направление ее деятельности определяются ни органом, ни анатомической близостью компонентов, а динамикой объединения, диктуемой только качеством конечного приспособительного эффекта.”

"Анохин П.К. А69 Избранные труды: Кибернетика функциональных систем/Под ред. К.В. Судакова"



# Взаимодействие - примеры функциональных систем

Сильносвязанные:

- Дыхание
- Система кровообращения
- и т.п.

Слабосвязанные:

- Муравейник
- Оркестр
- Организация
- и т.п.

# **Выбор** требуемого результата и **поиск** способа его достижения

Чем отличается машина от живого организма?

Организм динамически решает вопрос: “Что делать?” и затем “Как делать?”, а для машины вопрос “Что делать?” уже решён конструктором машины.

"Анохин П.К. А69 Избранные труды: Кибернетика функциональных систем/Под ред. К. В. Судакова"

# Саморегуляция

“Рецепторный аппарат гипоталамуса, оценивающий уровень осмотического давления крови, иногда на протяжении восьмидесяти лет остается абсолютно неизменным”.

"Анохин П.К. А69 Избранные труды: Кибернетика функциональных систем/Под ред. К. В. Судакова"

# Саморегуляция

# Подражание

Что делает живой попугай?

1. Слышит звук
2. Пытается издать звук, максимально похожий на услышанный

Что делает маленький человеческий ребёнок?

1. Слышит звук
2. Пытается издать звук, максимально похожий на услышанный

# Подражание

Что делает человек изучая чужой язык?

1. Слышит звук
2. Пытается издать звук, максимально похожий на услышанный

Что делает человек увидев жест или движение телом?

1. Видит движение
2. Пытается повторить движение

# Подражание

Что делает человек, изучая письменность?

1. Видит изображение буквы
2. Пытается повторить изображение буквы карандашом или ручкой на бумаге

Всё это один и тот же механизм - подражание.

Другими словами, универсальный механизм самообучения и обмена знаниями между живыми существами.

# Подражание, определение

Подражание — механизм социализации, следование образцу. У животных — подражательное (имитационное) научение, копирование поведения. За счёт подражания можно усвоить новые формы поведения, при этом подражание может вестись как на уровне самих воспроизводимых действий, так и на уровне осознания смысла этих действий. Подражательное поведение может быть бессознательным — к такому, например, относится «заразительное» зевание. Может быть как направленным, так и неосознанным.

За подражанием стоят разные психологические механизмы:

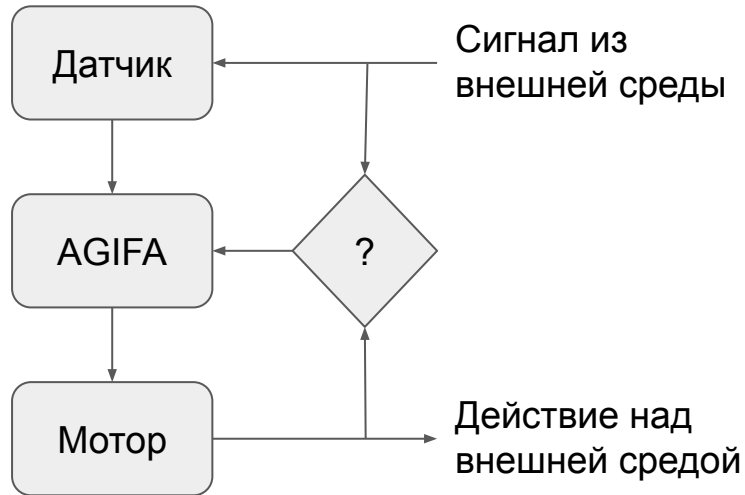
- в младенчестве — подражание движениям и звукам является попыткой установить контакт
- в детстве — проникновение в смысл человеческой деятельности через моделирование
- в игре
- в юности — идентификация с кумиром, принадлежность к группе
- в зрелом возрасте — научение в профессиональной деятельности

<https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B4%D1%80%D0%B0%D0%B6%D0%B0%D0%BD%D0%B8%D0%B5>



# Как сделать Подражание?

Схема использования Базового Компонента (БК) AGIFA в тестовой задаче Попугай



# Как сделать Подражание?

Датчик - это "ухо", т.е. введенный текст из консоли.

Мотор - это "мышца" (сжимается-разжимается), длина которой преобразуется в буквы, которые затем выводятся текстом в консоль.

1. Датчик получает из консоли ввода последовательно отдельные буквы текста
2. Ядро AGIFA автоматически строит *Дерево Результатов (ДР)*
3. Используя ДР, ядро AGIFA синтезирует управляющие команды на Мотор, в заданном рабочем диапазоне от 0 (мышца полностью расслаблена) до N (мышца сжата до максимума).
4. Текущее состояние Мотора преобразуется в соответствующие последовательности букв алфавита, которые выводятся в консоль вывода.

Пример:

Входные данные: “агу”

Выходные данные: “агу”

# Дерево Результатов (ДР), правила построения

1. ДР — это дерево, каждый узел которого состоит из Результата (название результата и параметры результата).
2. Каждый Результат может состоять из одного и более дочерних Результатов.
3. Каждый Результат имеет параметры «Дату-время начала» и «Длительность», т. е. узел ДР — это процесс во времени, по окончании выполнения которого получается требуемый результат.
4. Нельзя получить Родительский Результат, без достижения всех дочерних результатов.
5. Если название одного Результата совпадает с другим Результатом, то это считается одним и тем же результатом, т. е. один узел ДР может иметь более одного родительского узла.
6. Формулировка названия Результата должна отвечать на вопрос «Что?». Например, «Что требуется сделать?» - «Построить дом».
7. Дочерние результаты получаются посредством рекурсивного ответа на вопрос: «Что требуется сделать для достижения текущего Результата?».

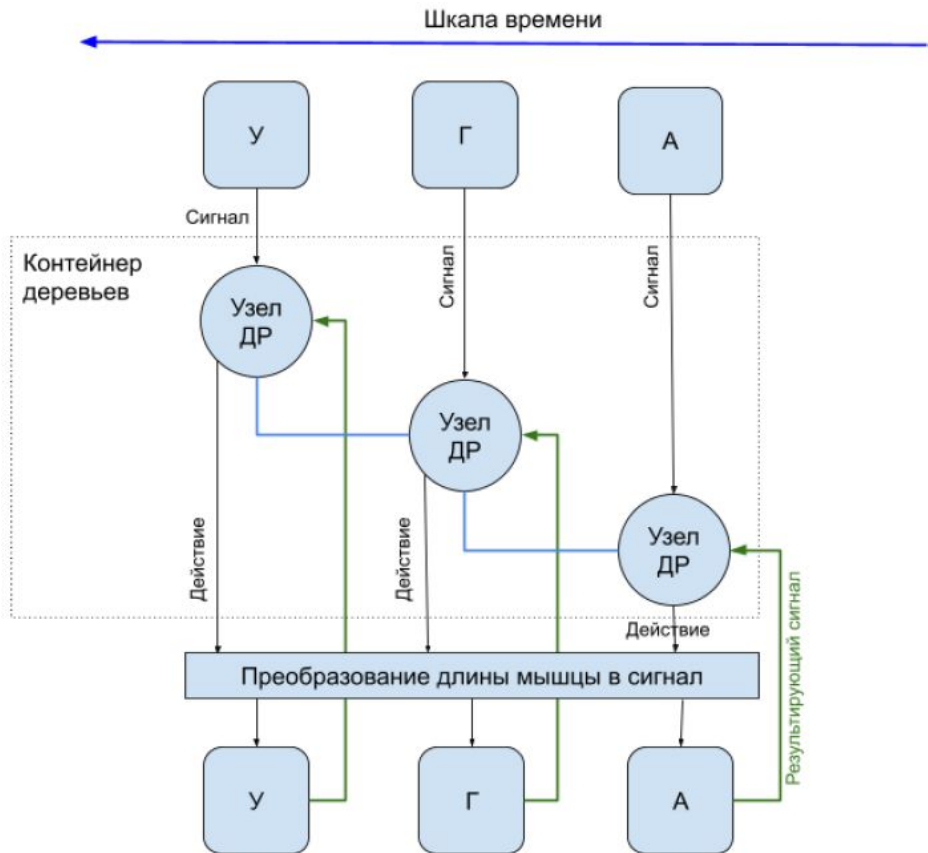
# Дерево Результатов (ДР), пример

- (5) Построить дом
  - (4) Сделать крышу
    - (3) Сделать стены
      - (2) Сделать фундамент
        - (1) Вырыть котлован
  - (4) Сделать двери
    - (3) Сделать стены
      - (2) Сделать фундамент
        - (1) Вырыть котлован *(это не дубль! А один узел)*
  - (4) Сделать окна
    - (3) Сделать стены
      - (2) Сделать фундамент *(аналогично, для всех совпадающих названий)*
        - (1) Вырыть котлован *(и это всё тот же узел)*

# Дерево Результатов (ДР), основные свойства

1. Узлы ДР проще сравнивать друг с другом, потому что все их дочерние элементы имеют строгие зависимости друг от друга. Например, нельзя построить дом без окон или без дверей.
2. ДР имеет строгую последовательность исполнения — от терминальных узлов вверх к родительским (см. нумерацию в скобках). Например, в примере строительство дома начинается строго с «Вырыть котлован» и завершается в корневом узле ДР «Построить дом».
3. В ДР сразу видны зависимые (последовательные) друг от друга процессы и независимые (параллельные). Например, результаты «Сделать окна», «Сделать двери», «Сделать крышу» можно достигать независимо (находятся на одном уровне ДР), а вот «Сделать стены» не получится раньше, чем «Сделать фундамент».

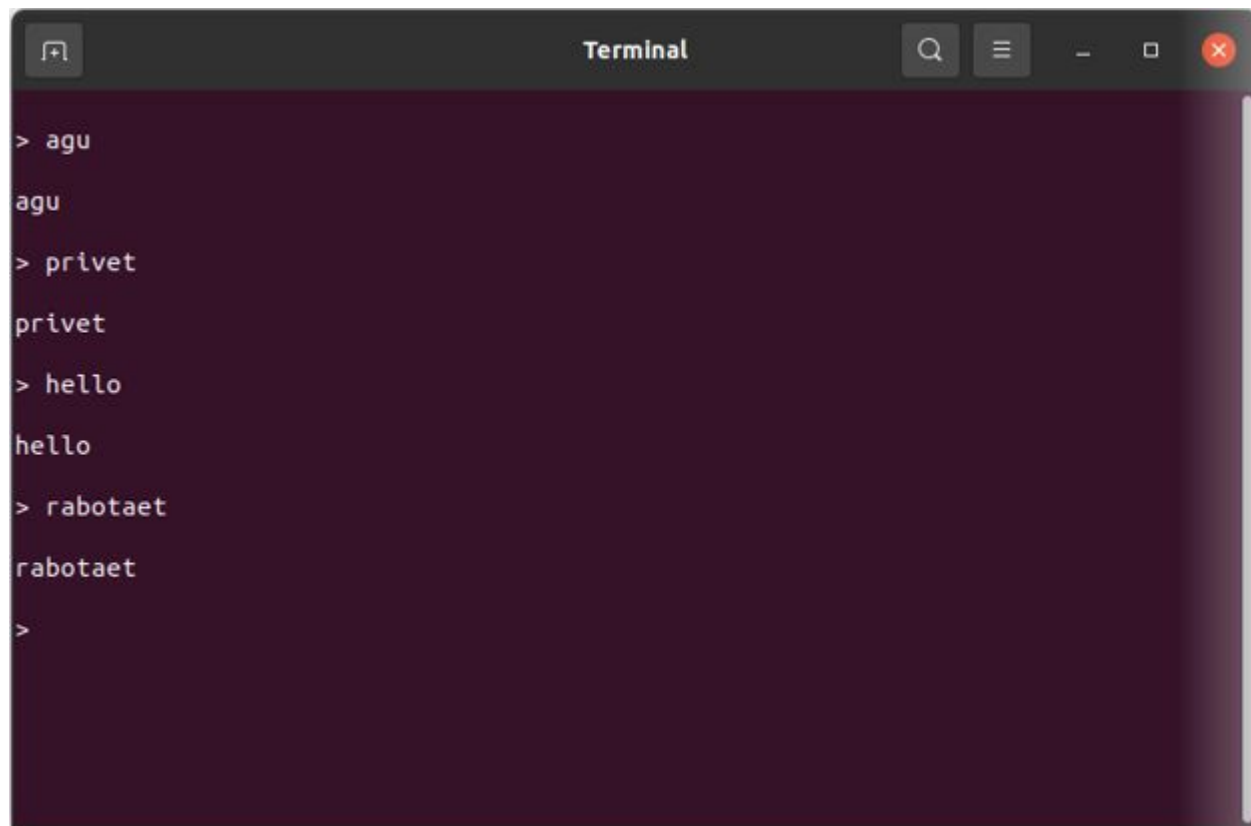
# Алгоритм подражания на основе Дерева Результатов (ДР)



# Алгоритм подражания на основе Дерева Результатов (ДР)

1. Каждый Узел ДР (УДР) имеет 2 входа: сигнал1 (С1), сигнал2 (С2) - результат действия (такой же сигнал с датчика, только с задержкой Т мс.) и один выход: действие (Д).
2. С1 и С2 сравниваются (распределённая реализация механизма Акцептора Действия).
3. Если С1 и С2 не равны, с заданной точностью, то УДР выполняет “ориентировочно-исследовательскую реакцию” (см. “4. АППАРАТ АКЦЕПТОРА ДЕЙСТВИЯ”) - синтезирует действия, которые сокращают мышцу от 0 до N.
4. Длина мышцы преобразуется в сигнал, который вызывает срабатывание датчика (“ухо” Попугая услышало изданный Попугаем “звук”) и таким образом в УДР поступает С2.
5. В какой-то момент времени, С1 становится равен С2. Сигнал С2 и крайнее выполненное действие Д запоминаются в УДР. В соответствии с правилом №4 ДР (“4. Нельзя получить Родительский Результат, без достижения всех дочерних результатов.”) происходит переключение УДР на вышестоящий родительский узел ДР. Если родительский узел пока ещё отсутствует, то создаётся новый УДР, в который уже на следующем такте поступает следующий сигнал С1 с датчика.

# Результат работы БК в тестовой задаче Попугай

A screenshot of a terminal window with a dark background. The window title is "Terminal". It shows a series of commands entered by a user and their corresponding outputs. The commands are "agu", "privet", "hello", and "rabotaet", each followed by its output on the next line. The prompt ">" is visible at the start of each command line and at the end of the last line.

```
> agu
agu
> privet
privet
> hello
hello
> rabotaet
rabotaet
>
```



У живых **Ухо** и **Голосовой аппарат** - разные компоненты!

Проблема:

Как из однокомпонентного Попугая сделать двухкомпонентного Попугая?

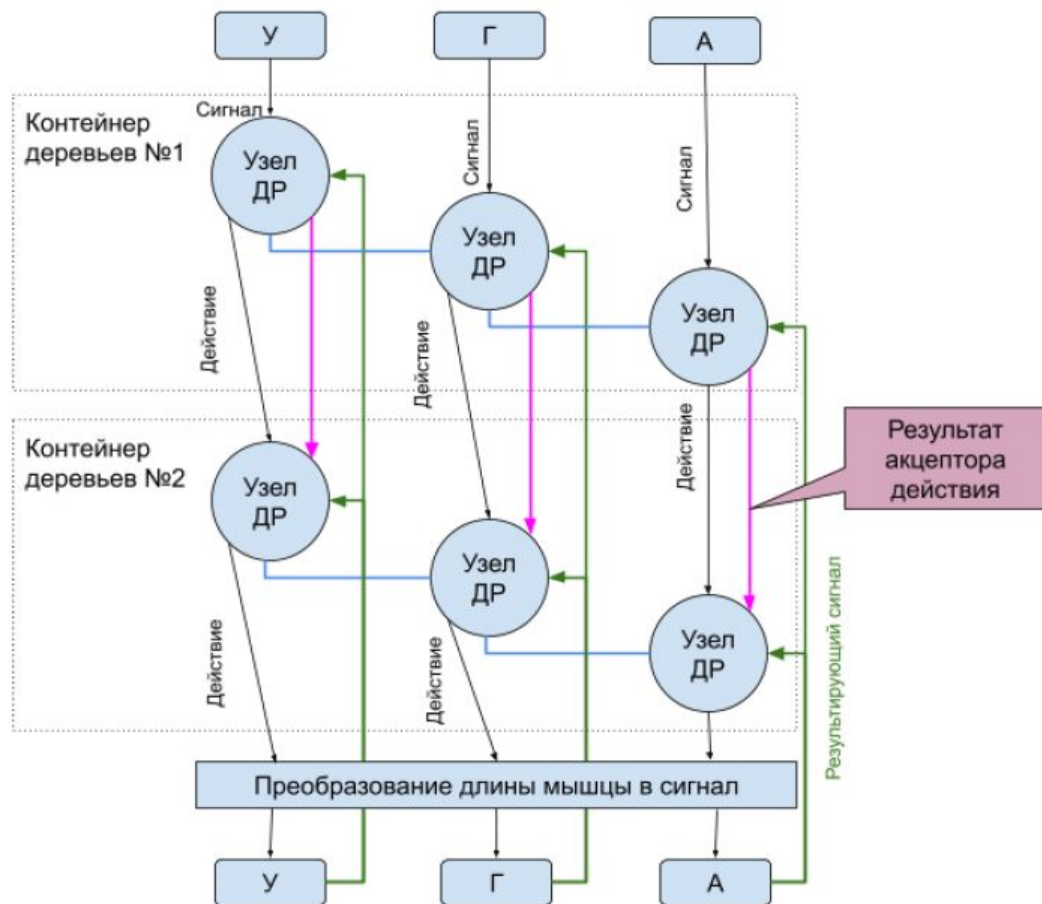
Решение:

Добавить второй компонент и как-то их заставить **взаимодействовать!**

Мысль вслух:

“Не может быть, чтобы так просто!”

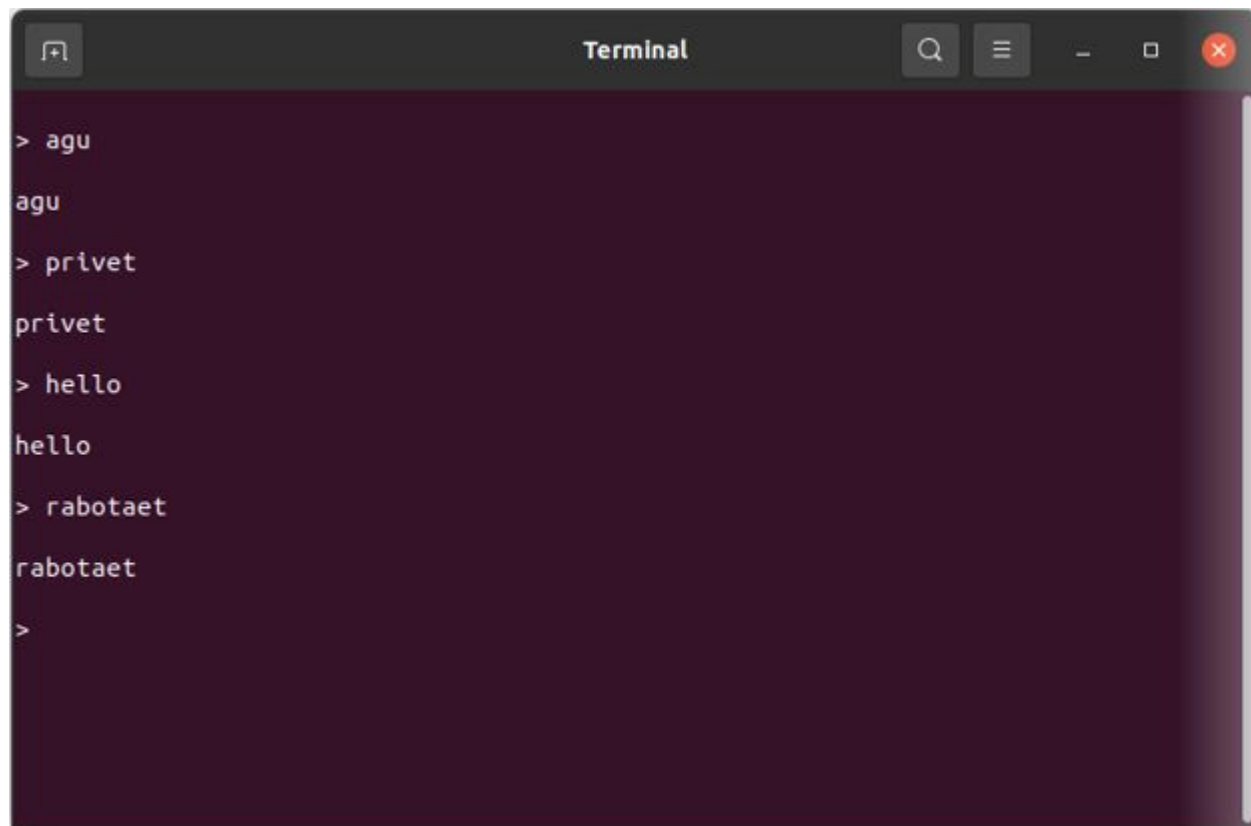
Шкала времени



# Двухкомпонентный алгоритм подражания Попугая

1. Каждый Узел ДР (УДР) имеет 2 или 3 входа: сигнал1 (С1), сигнал2 (С2) - результат действия (такой же сигнал с датчика, только с задержкой Т мс.), результат акцептора действия и один выход: действие (Д).
2. С1 и С2 сравниваются (распределённая реализация механизма Акцептора Действия), при наличии результата акцептора действия (разрешает сравнение, с целью синхронизации с другими компонентами).
3. Если С1 и С2 не равны в контейнере №1, с заданной точностью, то УДР выполняет “ориентировочно-исследовательскую реакцию” (см. “4. АППАРАТ АКЦЕПТОРА ДЕЙСТВИЯ”) - синтезирует действия, которые поступают на вход УДР контейнера №2.
4. Если С1 и С2 не равны в контейнере №2, с заданной точностью, то УДР выполняет “ориентировочно-исследовательскую реакцию” (см. “4. АППАРАТ АКЦЕПТОРА ДЕЙСТВИЯ”) - синтезирует действия, которые сокращают мышцу от 0 до N.
5. Длина мышцы преобразуется в сигнал, который вызывает срабатывание датчика (“ухо” Попугая услышало изданный Попугаем “звук”) и таким образом в УДР поступает С2.
6. В какой-то момент времени, С1 становится равен С2, в обоих контейнерах. Сигнал С2 и крайнее выполненное действие Д запоминаются в УДР. В соответствии с правилом №4 ДР (“4. Нельзя получить Родительский Результат, без достижения всех дочерних результатов.”) происходит переключение УДР на вышестоящий родительский узел ДР. Если родительский узел пока ещё отсутствует, то создаётся новый УДР, в который уже на следующем такте поступает следующий сигнал С1 с датчика.

# Результат работы взаимодействующих БК в задаче Попугай



```
> agu
agu
> privet
privet
> hello
hello
> rabotaet
rabotaet
>
```

The image shows a terminal window titled "Terminal" with a dark background and light text. It displays the output of a Parrot Game interaction between two bots. The first bot, "agu", sends the message "agu". The second bot, "privet", responds with "privet". The first bot then sends "hello", and the second bot responds with "hello". Finally, the first bot sends "rabotaet", and the second bot responds with "rabotaet". The terminal window has a standard macOS-style title bar with a search icon, a menu icon, and window control buttons (minimize, maximize, close).

# Результаты тестовой задачи Попугай

1. Реализован механизм подражания
2. Реализован механизм взаимодействия независимых компонентов в достижении общего требуемого результата

# У живых Ухо и Голосовой аппарат находятся на значительном удалении друг от друга

Вопрос:

Что может находиться между ними?

Ответ:

У живых - всё те же “нейроны”.

А в AGIFA если добавить третий БК и заставить его взаимодействовать с Ухом и Голосовым аппаратом, то промежуточный БК будет выполнять функцию передачи сигнала на удалённые друг от друга компоненты.

Схема из трёх БК пока *не реализована*, т.к. она требует параллельных компонентов - оказалось, что у текущих участников проекта, кроме меня, не хватает квалификации для написания соответствующего кода. Я пока почти всё в проекте делаю один, поэтому разработка продвигается медленно.

# Развитие проекта в ближайшем будущем

1. Реализовать параллельные БКР
2. Реализовать 3-х и более компонентную схему Попугая
3. Реализовать выбор требуемого результата и поиск способа его достижения
4. Реализовать тестовую задачу (пока без названия), демонстрирующую Саморегуляцию

# “Эффект” промежуточного БК

Предположим, что промежуточный 3-й БК добавили в схему Попугая.

Что полезного может появиться?

1. Живые умеют снижать активность разных отделов мозга, с целью отдыха. Если временно отключить БК Ухо, то промежуточный БК должен будет продолжать воспроизводить ранее услышанное и заставлять Голосовой Аппарат “бормотать”.
2. Если далее временно отключить Голосовой Аппарат, то промежуточный БК продолжит делать что? “Бормотать”. С точки зрения внешнего наблюдателя - это равносильно “думать про себя”.
3. Если обратно включить Голосовой Аппарат, то мы услышим, что “думает” Попугай в данный момент времени - он будет “бормотать вслух” или это равносильно “думать вслух”.



# Разговоры во сне

Разговоры во сне или сноговорение — это парасомническое расстройство, которое характеризуется разговором вслух во время сна, начиная от простых бормочущих звуков до громких криков и длинных, часто нечленораздельных речей. Может происходить много раз в течение цикла сна.

<https://ru.wikipedia.org>

Трёхкомпонентный Попугай демонстрирует свойства аналогичные наблюдаемым у человека. Если часть БК “спит”, а “промежуточные БК” уже “просыпаются”, то *будут наблюдаться эффекты аналогичные наблюдаемым у человека, включая **галлюцинации и психические расстройства*** (в этом случае часть БК повреждены).

# Условия возникновения мышления у живых

- достаточно большое расстояние между группами нейронов - функциональными системами
- способность отдельных функциональных систем впадать в сон, т.е. временно “отключаться”

При возникновении этих условий, сформировались промежуточные функциональные системы, которые сначала выполняли функцию передачи сигналов на расстояние, а затем в них возникли процессы, при временном “отключении” крайних групп нейронов, которые мы сейчас идентифицируем как “мышление”, “воображение” и т.п..

# Как “бормотание” становится “мышлением”?

Предлагаю подумать над этим пока самостоятельно - решение на “самом видном месте”.

**Продолжение будет** после реализации перечисленных в начала 4-х требований к БКР в AGIFA.

# Вопросы и ответы

1. В каком виде ставится задача?

Задача ставится в виде многомерного образа требуемого результата, т.е. показания датчиков в момент получения требуемого результата.

2. Бывает ли ситуация, что формулировка существует, а агент не может достичь этой цели?

Бывает.

3. Бывает ли ситуация, что формулировка существует, агент не может достичь цели, но некий рукописный алгоритм - может?

Бывает. Зависит от полноты знаний о мире.

# Вопросы и ответы

1. Насколько система может работать, не имея на старте знаний о среде?

Система изначально спроектирована, чтобы так работать.

2. Система может научиться проходить лабиринт, если лично вы не знаете его карту?

Потенциально может.

3. У вас же программа получается детерминированной, а значит, она всегда будет поворачивать или влево, или вправо, или глядя на стенки, но все равно одинаковым образом.

Это утверждение не верно.

**Спасибо за внимание!**

Артюхов Виктор (@LiveBaster)

<https://github.com/LiveBaster/agifa>