РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Материалы к обсуждению «Искусственный интеллект: онтологический и социальный аспекты

Докладчик

Зав.каф. когнитивно-аналитических и нейро-прикладных технологий,

д.т.н., проф. Щербаков А.Ю.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Что мы хотим обсудить?

- Онтология искусственного интеллекта (ИИ) – что такое ИИ как феномен?

- Ключевой момент – ИИ в полной мере еще не существует, есть только элементы технологии, то есть, в первую очередь мы говорим о «прогностическом дискурсе» - «Может ли машина мыслить?» и «Что это дает человеку и социуму?»

- Социальный аспект – сферы применения ИИ, взаимодействие человека и ИИ и как может измениться будущее с появлением ИИ?

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Текущее «формальное состояние» проблемы

- Искусственный интеллект (ИИ) – комплекс технологических решений, позволяющий **имитировать** когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

- Комплекс технологических решений (КТР) включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Текущее «формальное состояние» проблемы

- Технологии ИИ – технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и **перспективные методы искусственного интеллекта.**

- Технологические решения – технология, программа для электронно-вычислительных машин (программа для ЭВМ), база данных или их совокупность, а также сведения о наиболее эффективных способах их использования.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Что относится к ИИ?

1. Компьютерное зрение
Примеры: распознавание дефектов продукции на основе анализа различных типов изображений; обнаружение и идентификация людей и объектов в сложной окружающей среде; генерация и анализ изображений и видеозаписей.

2. Обработка естественного языка
Примеры: машинный перевод; классификация и кластеризация отдельных высказываний, коротких и длинных текстов; извлечение фактов из текста и их систематизация.

3. Распознавание и синтез речи
Примеры: создание мультизадачных разговорных ассистентов; распознавание звуков и речи в сложных условиях (шумы, большое расстояние и т.д.), синтез речи на иностранном языке.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Что относится к ИИ?

4. Интеллектуальная поддержка принятия решений
Примеры: предиктивный и прескриптивный анализ, позволяющий предсказывать развитие ситуации на основе анализа данных и автоматизировать принятие решений в режиме реального времени; оценка качества моделей машинного обучения без тестирования в реальной среде; прогноз качества выпускаемой продукции, в частности прогноз вероятности и типов дефектов продукции, в том числе позволяющий находить и устранять причины этих дефектов.

5. Перспективные методы ИИ
Примеры: автономное решение различных задач; автоматический дизайн физических объектов; автоматическое машинное обучение; алгоритмы решения задач на основе данных с частичной разметкой и/ или незначительных объемов данных; **обработка информации на основе новых типов вычислительных систем;** интерпретируемая обработка данных.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Текущее «формальное состояние» Обсуждение

Официальный дискурс говорит только об **имитации** человеческих когнитивных процессов.

Имитация тесно связана с реализацией нейросетей как технического воплощения структуры и функций человеческого мозга.

Этот подход «тактически» конструктивен (дает на текущем этапе много статистически удачных результатов для общественной позитивной оценки), но:

- сужает область научного поиска,
- исключает альтернативные когнитивные процессы,
- практически исключает анализ проблемы содержания и смыслов,
- минимизирует возможность объяснений действий ИИ (связано с предыдущим пунктом).

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Исходная точка

Может ли машина мыслить? (60-70 гг. XX века)

Искусственный интеллект – субъект, реализующий помощь человеку в процессе автоматизации интеллектуальных процессов («помощник» и «собеседник»).

Искусственный разум – субъект, осознающий свое Я и самостоятельно ведущий мыслительные (когнитивные) процессы (равный или превосходящий человека).

Сейчас категория «искусственный разум» заменена на «сильный ИИ».

Основной тезис – ИИ субъектен.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Основные подходы к ИИ

1. Картезианский подход.

Объединяет идеи, излагаемые в работах Р. Декарта, Н. Мальбранша, Дж. Локка, Б. Спинозы, Г. Лейбница. Характеризуется опорой на картезианскую методологию, основанную на принципе **самоочевидности исходных посылок**. В современной философии сознания его значимость и влияние распространяется на область мысленных экспериментов, основывающихся на апелляции к здравому смыслу, называемых картезианскими аргументами («аргумент китайской комнаты», «китайская нация», парадокс феноменальных суждений и т.д.).

2. Субъективистский подход.

Основные идеи выражены в работах Дж. Беркли, Д. Юма, И. Канта, И. Г. Фихте. Предполагает активную роль субъекта в осознании феноменального мира (мира как совокупности ощущаемых и воспринимаемых субъектом феноменов).

3. Натуралистический (биологический) подход.

Развивается в работах философов Просвещения и большинства современных философов (Дж. Сёрла, Д. Чалмерса, Т. Нагеля и др.). Положением, лежащем в его основе является тезис о производности ментального от физического. Также, данный подход подразумевает необходимость согласования концепции сознания с научной картиной мира.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Основные подходы к ИИ

4. Объективирующий подход.

Объединяет идеи, развиваемые в рамках таких направлений мысли как бихевиоризм, физикализм, элиминативный материализм. Отличительной особенностью его сторонников является стремление к рассмотрению сознания с позиций третьего лица и к преодолению субъективного характера сознания путём его описания в терминах поведения (с точки зрения информатики – в терминах свойств и функций субъекта).

5. Функционалистский подход.

Отражён в работах Х. Патнема, Д. Деннета и ряда других философов. Характерной особенностью является утверждение в качестве основополагающего тезиса о достаточности идентичности системы для изоморфности продуцируемых субъективных реальностей, а также приверженность тезису об организационной инвариантности системы.

6. Информационный подход.

Развивается в работах Д. И. Дубровского, А. Г. Спиркина, Д. Чалмерса. Характеризуется привлечением для описания сознательных феноменов терминологии теории информации (в частности, понятий информация, информационная причинность и т.д.).

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Пример. Мысленный эксперимент «Китайская комната»

Эксперимент был опубликован в 1980 году в статье «Minds, Brains, and Programs» журнала «The Behavioral and Brain Sciences ». Ещё до публикации эксперимент вызвал полемику, поэтому статья содержала как оригинальный аргумент Дж.Сёрла, так и возражения многих исследователей в области когнитивных наук, а также ответы Сёрла на эти возражения. Помимо этого, эксперимент описан в книге 1984 года «Minds, Brains and Science» и в январском выпуске журнала Scientific American 1990 года.

Описание эксперимента

Представим себе изолированную комнату, в которой находится Джон Сёрл, который не знает ни одного китайского иероглифа. Однако у него есть точные инструкции по манипуляции иероглифами вида «Возьмите такой-то иероглиф из корзинки номер один и поместите его рядом с таким-то иероглифом из корзинки номер два», но в этих инструкциях отсутствует информация о значении этих иероглифов, и Сёрл просто следует этим инструкциям подобно компьютеру.

Наблюдатель, знающий китайские иероглифы, через щель передаёт в комнату иероглифы с вопросами, а на выходе ожидает получить осознанный ответ. Инструкция же составлена таким образом, что после применения всех шагов к иероглифам вопроса они преобразуются в иероглифы ответа.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Пример. Мысленный эксперимент «Китайская комната»

Фактически инструкция — это подобие компьютерного алгоритма, а Сёрл исполняет алгоритм так же, как его исполнил бы компьютер.

В такой ситуации наблюдатель может отправить в комнату любой осмысленный вопрос (например, «Какой цвет вам больше всего нравится?») и получить на него осмысленный ответ (например, «Синий»), как при разговоре с человеком, который свободно владеет китайской письменностью. При этом сам Сёрл не имеет никаких знаний об иероглифах и не может научиться ими пользоваться, поскольку не может узнать значение даже одного символа. Сёрл не понимает ни изначального вопроса, ни ответа, который сам составил. Наблюдатель, в свою очередь, может быть убежден в том, что в комнате находится человек, который знает и понимает иероглифы.

Аргументация

Таким образом Сёрл заключает, что хотя такая система и может пройти тест Тьюринга, но при этом никакого понимания языка внутри системы не происходит, а значит тест Тьюринга не является адекватной проверкой мыслительных способностей. Доводы Сёрла направлены на критику позиции так называемого «сильного» искусственного интеллекта, согласно которой компьютеры с соответствующей программой на самом деле могут понимать естественный язык, а также обладать другими ментальными способностями, свойственными людям.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Ключевой тезис

Ключевой тезис когнитивной науки, называемый **компьютерной метафорой сознания**.

Данный тезис гласит, что сознание относится к мозгу таким же образом, как компьютерная программа относится к вычислительному устройству, в котором она исполняется.

Но в мысленном эксперименте «Китайская комната» мы видим неявный отсыл к семантическим процессам, т.е. полагается, что ИИ управляет процессами преобразования и формирования текстов.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Субъектно-объектная модель

Что мы упускаем?

Важный аргумент, связанный с поверхностным знанием философов о работе вычислительной техники.

Философия бытия и познания работает с точки зрения техника (программиста) с понятием типа «поток» (поток информации или поток, изменяющий материальные объекты).

Stream(Si, Oj)-> Om – поток информации от объекта Oj к объекту Om, реализуемый (управляемый) субъектом Si.

Например, пользователь при помощи редактора Word (субъект) изменяет текст Oj так, что из него получается текст Om.

В «Китайской комнате» мы работаем с текстами и реализуем процесс ответа на вопросы.

«От живого созерцания к абстрактному мышлению, и от него – к практике» (В.И. Ленин)

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Субъектно-объектная модель

В реальном мире мы имеем дело еще с одним «отображением» - порождение субъекта («творение»)

Create(Si, Oj)-> Sm – субъект Sj порождает новый субъект Sm из объекта-источника Oj

Пример: запуск новой программы из пассивного исполняемого файла.

Также полностью соответствует «креативной» и «эволюционной» картине мира («Большой взрыв» и последующее усложнение материи).

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Субъектно-объектная модель

Таким образом, первоначально созданная для нужд компьютерной безопасности модель становится конструктивным инструментом не только описания мира, но и синтеза новых сущностей, в том числе и для ИИ.

В первую очередь, это «креативизация» вычислительного процесса, когда вычислитель сам генерирует код (объект-источник) для дальнейшего использования, т.е. сам изменяет и в идеале «совершенствует» свое функционирование.

Кроме того, расширение ИИ позволяет «отвязаться» от ложной сервильности сильного ИИ, который уже не обязан быть полезным человеку, а представляет из себя самодостаточную и самостоятельную сущность.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Где же семантика?

«Вначале было слово. И слово было два байта…»

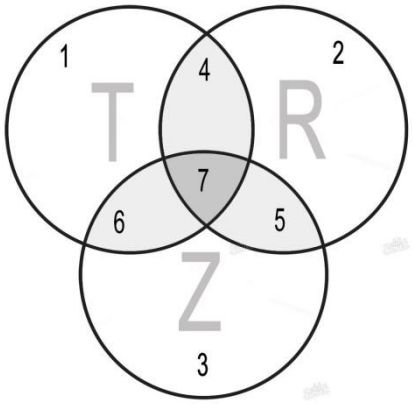
Мы общаемся друг с другом посредством слов и текстов, визуальный ряд тоже описывается словами (сценарий), но он воспринимается уже не так однозначно за счет того, что в голове у каждого Homo Sapiens своя настроенная и обученная нейросеть.

Но вполне возможно конструктивно и однозначно общаться словами и текстами.

И именно с этого необходимо начинать создание реального ИИ.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Мир, сознание и подсознание когнитивной системы как семантические объекты



Пусть T – это информация от окружающего мира, выраженная в виде слов-понятий, R – сознание мыслящей (когнитивной) системы (КоС) виде полного на текущий момент набора понятий (слов, объектов), осмысленных (имеющихся в распоряжении КоС), Z – подсознание (понимаемое как область интуитивных восприятий), также выраженное в понятиях (словах, выражениях).

Тогда возможны следующие интерпретации множеств:

1 – информация окружающего мира, не воспринятая когнитивной системой,

2 – набор понятий сознания, не участвующих в процессах восприятия (пассивные знания и навыки),

3 – подсознательная область, пассивная на данный момент.

4 – область соприкосновения окружающего мира и сознания (например, вся совокупность образов, поданная сознанию или воспринятая сознанием от органов чувств).

5 – область влияния бессознательного в сознании,

6 – область интерпретации окружающего мира подсознанием,

7 – точка текущего восприятия (мысль, точка сборки или фокус сознания) КоС.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Социальное измерение ИИ

Три закона робототехники Айзека Азимова
«Робот не может причинить вред человеку»

Парадокс Карениной
На одной ветке железнодорожного пути лежит Анна Каренина, на другой – стоит состав с тысячей пассажиров.

Сильный ИИ должен выбрать, куда направить паровоз – на ветку с Карениной или на ветку с составом?

С учетом, возможно, того фактора, что А.К. добровольно собралась уйти в лучший мир.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Социальное измерение (Кодекс этики в сфере ИИ)

1. Человеко-ориентированный и гуманистический подход. При развитии технологий ИИ человек, его права и свободы должны рассматриваться как наивысшая ценность. Разрабатываемые Акторами технологии ИИ должны способствовать или не препятствовать реализации всех потенциальных возможностей человека для достижения гармонии в социальной, экономической, духовной сфере и наивысшего расцвета личности, учитывать ключевые ценности, такие как: сохранение и развитие когнитивных способностей человека и его творческого потенциала; сохранение нравственных, духовных и культурных ценностей; содействие культурному и языковому многообразию, самобытности; сохранение традиций и устоев наций, народов, этносов и социальных групп.

2. Уважение автономии и свободы воли человека. Акторы ИИ должны принимать необходимые меры, направленные на сохранение автономии и свободы воли человека в принятии им решений, права выбора и в целом сохранения интеллектуальных способностей человека как самостоятельной ценности и системообразующего фактора современной цивилизации. Акторы ИИ должны на этапе создания СИИ прогнозировать возможные негативные последствия для развития когнитивных способностей человека, и не допускать разработку СИИ, которые целенаправленно вызывают такие последствия.

3. Соответствие закону. Акторы ИИ должны знать и соблюдать положения законодательства во всех сферах своей деятельности и на всех этапах создания, внедрения и использования технологий ИИ, в том числе в вопросах юридической ответственности Акторов.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ
РГСУ

Социальное измерение (Кодекс этики в сфере ИИ)

4. Недискриминация. В целях обеспечения справедливости и недопущения дискриминации Акторы ИИ должны принимать меры для того, чтобы удостовериться, что применяемые ими алгоритмы и наборы данных, методы обработки используемых для машинного обучения данных, при помощи которых осуществляется группирование и/или классификация данных, касающихся отдельных лиц или групп лиц, не влекут их умышленную дискриминацию.

5. Оценка рисков и гуманитарного воздействия. Акторам ИИ рекомендуется проводить оценку потенциальных рисков применения СИИ, включая социальные последствия для человека, общества и государства, гуманитарного воздействия СИИ на права и свободы человека на разных стадиях ее жизненного цикла, в том числе при формировании и использовании наборов данных; осуществлять долгосрочный мониторинг проявления таких рисков; учитывать сложность поведения СИИ, включая взаимосвязь и взаимозависимость процессов в жизненном цикле СИИ при оценке рисков. Для критических приложений СИИ в особых случаях приветствуется проведение оценки рисков посредством привлечения нейтральной третьей стороны или уполномоченного официального органа.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ СОЦИАЛЬНЫЙ УНИВЕРСИТЕТ

РГСУ

Благодарю за внимание!