

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**

**Белорусский национальный технический университет**

---

Факультет технологий управления и гуманитаризации

Кафедра «Философские учения»

**А.И. Лойко**

**ТЕОРИЯ РАЗУМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

Учебное пособие по общеобразовательной дисциплине

«Философия и методология науки»

для магистрантов, слушателей, осваивающих содержание образовательной  
программы магистратуры

Электронный учебный материал

Минск

БНТУ

2025

**Автор:**

А. И. Лойко, заведующий кафедрой «Философские учения» БНТУ, доктор философских наук, профессор

**Рецензенты:**

Барсук И.А., доцент кафедры философии и идеологической работы научно-педагогического факультета учреждения образования «Академия Министерства внутренних дел Республики Беларусь» кандидат философских наук доцент.

Кандричина И. Н., заведующий кафедрой «Менеджмент» Белорусского национального технического университета, кандидат социологических наук, доцент.

Пособие знакомит студентов с ролью философии в профессиональной деятельности. Оно состоит из разделов. В них изложены особенности прикладной функции философии. Важная роль отведена прикладным разделам, поскольку они связаны с анализом технологий искусственного интеллекта, четвертой промышленной революции, кибернетической безопасности, цифровой антропологии, цифровой этики и философии технологий цифрового дизайна.

Электронное учебное пособие может быть использовано для организации самостоятельной и научно-исследовательской работы студентов.

© Лойко А.И. 2025

© Белорусский национальный  
технический университет, 2025

## **ВВЕДЕНИЕ**

В пособии излагаются прикладные аспекты теории разума искусственного интеллекта. Необходимость анализа прикладных аспектов рассматриваемой теории обусловлена неопределённостью в определениях терминов и высокими рисками подмены понятий, что может сказаться на привлекательности развития технологий искусственного интеллекта. Это важно для будущих сотрудников крупных компаний и банков, которые оперируют большими объёмами данных и доверяют агентам искусственного интеллекта автономное решение организационных и управленческих задач. Поэтому, чтобы адекватно судить об уровне интеллектуальных возможностей технологий искусственного интеллекта важно построить концептуальную структуру проблемы и придать ей статус предметной онтологии, которая будет рассматриваться в функции критерия оценки реальных возможностей компьютерных когнитивных моделей, операционных систем и аппаратной части устройств диалоговых технологий искусственного интеллекта.

Электронное учебное пособие разработано на основе типовой программы по философии и методологии, утвержденной Министерством образования Республики Беларусь в 2022 г. Оно содержит лекционный материал и список учебной литературы для самостоятельной подготовки студентов.

### **Теория разума искусственного интеллекта агентов**

Эта теория предписывает разработчикам наделять технологии ИИ способностью наблюдения за поведением и общением людей. Технология изучает особенности взаимодействия и собирает данные о различных моделях поведения людей и контекстах. На основе этой информации технология ИИ может распознавать закономерности в мыслях и чувствах людей. У технологии ИИ формируется перспектива использовать семантику мышления и чувств человека в той или иной ситуации. Технология ИИ корректирует неточности и меняет подход к решению ситуации.

Если пользователь отвечает кратко на вопросы, то технология ИИ переходит на режим диалога коротких и прямых ответов. Теория разума ИИ объясняет, как распознавать эмоции человека и выражения и преобразовывать их в компьютерные модели. Системы искусственного интеллекта могут распознавать эмоции человека по выражению лица, тону голоса и анализу тональности текста.

Имея в своем распоряжении исторические наборы данных, прогностические алгоритмы, основанные на машинном обучении, могут предсказать, что конкретный человек может сделать в следующий раз. Рекомендательные системы предлагают продукты, фильмы и музыку на основе прошлых выборов пользователя.

Чат боты и виртуальные помощники, стали улавливать контекст, намерения и эмоциональное состояние пользователей сети через использование сложных диалоговых систем. Это делает взаимодействие с пользователем естественным, поскольку оно кажется эмпатичным.

Имитируя процессы мышления человека, модели ИИ могут предсказывать поведение пользователя в различных обстоятельствах. Многоагентные системы предвидят действия и намерения, в которых участвуют другие ИИ.

Для решения технологических задач важно понять, как люди воспринимают эмоции и убеждения. Не все люди воспринимают ситуации одинаково. Системы ИИ наделяются способностью интерпретировать вербальные и невербальные сигналы. Дифференциация этих сигналов может быть сложной, учитывая широкий диапазон уровней эмоциональной культуры людей.

Актуальным направлением является разработка ментальных моделей. Это предполагает наличие точных представлений о других интеллектуальных существах, представленных людьми и социальными роботами.

Обучение обучению, предполагает анализ эффективности различных моделей машинного обучения при решении конкретных задач. Данный подход позволяет технологии ИИ использовать знания или данные для эффективного решения задач.

Реализуются программы создания агентов ИИ с машинной теорией разума, позволяющих им эффективно взаимодействовать в командной среде, наблюдая окружающую среду и партнёров людей.

Хотя программа находится на начальном этапе, есть результаты в виде распознавания эмоций, прогностического моделирования, разговорного ИИ, когнитивного моделирования и много агентных систем. Способность точно реагировать на мысли человека и эмоции известна как теория разума (ТоМ). Она может влиять на то, как люди реагируют на рекомендации ИИ.

Достижения в технологии искусственного интеллекта отмечены значительными инновациями, такими как усовершенствованная модель ИИ DeepSeek (2025), чип Blackwell (NVIDIA 2025) и введение Grok 3 (xAI 2025). Прогресс в развитии технологий человекоподобного ИИ меняет способ взаимодействия компаний с потребителями. Эмоциональный ИИ способствует повышению антропологии ИИ, распознавая эмоции и реагируя на них в режиме реального времени.

Пользователи ожидают от ИИ демонстрации определённого уровня социального интеллекта, адаптивности и чувствительности к их потребностям. Центральной конструкцией, отражающей эту способность, является теория разума (ТоМ). Она относится к способности ИИ точно определять ментальные состояния, мысли и намерения пользователя, что позволяет ему реагировать эмпатично и адекватно с точки зрения пользователя.

Во взаимодействиях человека и ИИ, связанных с услугами, агенты ИИ обычно рассматриваются либо как партнеры, либо как слуги. ИИ в ролях партнеров делает акцент на руководстве и поддержке долгосрочных целей. Те, кто играет роли слуг, фокусируются на инструментальной полезности и уважении к автономии пользователей. Различное восприятие ролей может зависеть от типа продукта добродетели или порока.

Продукты добродетели подразумевают краткосрочные затраты в погоне за долгосрочными выгодами. Продукты порока предлагают немедленное удовольствие, но могут подорвать долгосрочные затраты. При выборе,

связанном с добродетелью, потребители могут ожидать от ИИ поддержки, ориентированной на достижение цели. В случае с пороком они могут предпочесть, чтобы ИИ оставался пассивным и непредвзятым.

Теория нарушения ожиданий предполагает, что люди формируют поведенческие ожидания на основе контекстных норм. Они реагируют в зависимости от того, нарушаются ли эти ожидания. ИИ с высоким уровнем ТоМ повышает принятие рекомендаций, превосходя ожидания пользователей и усиливая воспринимаемое социальное присутствие.

ИИ с высоким уровнем ТоМ может вызывать позитивные нарушения ожиданий в контексте добродетели, но провоцировать негативные в контексте порока из-за попыток направлять или влиять на выбор пользователя. ИИ с низким уровнем ТоМ может лучше соответствовать ожиданиям потребителей в контексте порока, воздерживаясь от вынесения нормативных суждений и вместо этого пассивно выполняя команды пользователя.

Эмоциональный ИИ вводит новое измерение во взаимодействие человека и ИИ. Он демонстрирует интегрированные когнитивные и эмоциональные функции, включая имитируемую эмпатию, автономность и личность. Это отличает его от механического, аналитического и интуитивного ИИ. Возникли этические вопросы, касающиеся человеческой идентичности, автономии и конфиденциальности

Без способности предугадывать предпочтения пользователей или адаптироваться к тонким контекстным подсказкам ИИ с низким ТоМ может генерировать рекомендации, которые кажутся общими, нерелевантными или даже навязчивыми. Пользователи могут испытывать разочарование, когда системы ИИ неоднократно предлагают продукты, которые не соответствуют их намерениям.

Такие функции ИИ, как тон голоса, выражения лица и гендерно-специфические сигналы, могут активировать ментальные схемы пользователей, связанные с человеческим общением. Пользователи применяют социальные нормы и ожидания во время этих обменов. Это усиливает человеко-

подобные качества агентов, заставляя пользователей взаимодействовать с ними способами, которые отражают человеческие взаимодействия.

Когда пользователи сети воспринимают агента ИИ как социально присутствующего, они с большей вероятностью будут рассматривать его рекомендации как продуманные предложения от социально осведомленной сущности, а не просто механические результаты. Когда пользователи чувствуют, что агент ИИ «понимает» их эмоциональное и когнитивное состояние, они с большей вероятностью будут воспринимать взаимодействие как личное и искреннее и принимать предложения ИИ. В результате социальное присутствие действует как посредник между возможностями ТоМ агента ИИ и принятием пользователем его рекомендаций.

Люди формируют ожидания относительно поведения других людей на основе социальных норм, контекстных сигналов и прошлого опыта. Эти ожидания служат точкой отсчета для интерпретации того, соответствует ли коммуникативный акт тому, что считается уместным в данной ситуации. В контексте взаимодействия человека и ИИ такие ожидания могут распространяться на роль, которую, как ожидается, будет играть ИИ, в зависимости от основных целей потребителя или мотивации различных контекстов продукта.

Теория разума относится к способности человека рассуждать о ментальном содержании других людей, таком как их убеждения, желания и цели. Люди используют свою теорию разума для понимания, рассуждения и объяснения поведения других. Наличие теории разума особенно полезно при сотрудничестве людей, поскольку в этом случае люди могут рассуждать о том, что знает другой человек, а также о том, какие рассуждения он мог бы использовать. Аналогичным образом, гибридные интеллектуальные системы, в которых агенты ИИ сотрудничают с людьми, требуют, чтобы агенты рассуждали о людях, используя вычислительную теорию разума. Однако попытка отслеживать все индивидуальные ментальные установки всех других людей становится вычислительно очень сложной.

Важную роль играет механизм вычислительной теории разума, основанный на абстракциях отдельных убеждений в более высокоуровневые концепции. Эти абстракции могут быть вызваны социальными нормами и ролями. Их использование в процессе принятия решений служит эвристикой для выбора между взаимодействиями, тем самым способствуя сотрудничеству. Используется формализация, основанная на эпистемической логике, чтобы объяснить, как различные выводы позволяют создать такую вычислительную теорию разума.

### **Экосистема агентов искусственного интеллекта**

Агенты ИИ являются автономными сущностями, которые воспринимают, обрабатывают информацию и действуют для достижения определённых целей, варьируясь от простых реактивных систем до сложных самообучающихся систем. Их значение заключается в автоматизации задач, повышении эффективности и совершенствовании процесса принятия решений в различных отраслях, включая финансы и здравоохранение.

Понимание агентов ИИ необходимо для эффективного применения, управления и этичного использования, обеспечения ответственного развития и снижения рисков. Они оказывают значительное влияние на отрасли, оптимизируя процессы, персонализируя взаимодействие и упрощая принятие решений на основе данных.

Реактивные агенты ИИ воспринимают окружающую среду и реагируют непосредственно на стимулы, не обладая внутренним состоянием или памятью. Их интеллект основан исключительно на способности реагировать на текущие входные данные таким образом, чтобы соответствовать их замыслу, но они не могут хранить или использовать прошлый опыт для формирования будущих действий.

Реактивные агенты не хранят предыдущие взаимодействия или события. Они принимают решения, основываясь на текущей среде. Агенты ис-

пользуют predetermined правила или эвристики, чтобы решить, как реагировать на входные данные, что делает их детерминированными.

Их действия привязаны к определенным входящим данным, что означает, что они не могут адаптироваться к новым ситуациям, если их не перепрограммировать.

Базовые чат-боты, основанные на правилах, предоставляют предустановленные ответы на основе запросов пользователей, но не могут подстраиваться под разговоры, выходящие за рамки их predetermined рамок. NPC в видеоиграх действуют как реактивные агенты, следуя набору predetermined шаблонов или моделей поведения в ответ на действия игрока.

Поскольку реактивные агенты могут мгновенно реагировать на входные данные, они идеально подходят для приложений, требующих быстрых, заранее определенных действий без сложных рассуждений.

Чат-боты: используются в службах поддержки клиентов или ответах на часто задаваемые вопросы для предоставления мгновенных ответов.

Игровые NPC улучшают игровой процесс. Они динамически реагируют на действия игроков. К рассматриваемому классу технологий ИИ относятся датчики освещенности или регуляторы температуры.

Реактивные агенты обеспечивают быструю реакцию на стимулы в режиме реального времени, что делает их идеальными для приложений, где задержка имеет решающее значение. Их легко проектировать и реализовывать, поскольку они следуют простым правилам. Простота использования снижает вероятность непредвиденного поведения и ошибок.

Поскольку они не могут учиться на прошлом опыте, они не могут со временем совершенствоваться или адаптироваться к новым условиям. Реактивные агенты связаны своими предустановленными правилами, что делает их менее эффективными в сложных или изменяющихся условиях.

Агенты с ограниченной памятью могут хранить и использовать прошлую информацию для принятия текущих решений. Эти агенты объединяют данные, поступающие в режиме реального времени, с историческими данными.

ми, чтобы со временем улучшить свою эффективность и качество принятия решений. Хотя они не способны к саморефлексии и пониманию эмоций, их способность помнить прошлые действия позволяет им принимать более обоснованные и точные решения.

Агенты с ограниченной памятью используют исторические данные или кратковременную память для анализа закономерностей и тенденций. Эта способность позволяет им учиться на основе последовательности действий и результатов, корректировать будущие ответы на основе предыдущего опыта или изменений окружающей среды, повышать производительность с помощью итеративных процессов машинного обучения.

Автономные транспортные средства используют искусственный интеллект с ограниченной памятью для оценки дорожной ситуации в режиме реального времени, дорожных препятствий и схем движения на основе данных о прошлых поездках.

Они принимают решения, такие как регулировка скорости, поддержание безопасной дистанции и объезд опасностей, используя сохранённые данные о предыдущих взаимодействиях с аналогичными ситуациями.

Виртуальные помощники используют ограниченную память для запоминания пользовательских предпочтений, предыдущих запросов и часто используемой информации, чтобы предоставлять более качественные и персонализированные ответы.

Опираясь на прошлые данные, агенты с ограниченной памятью могут принимать более точные и актуальные решения по сравнению с реактивными агентами. Они могут анализировать закономерности и тенденции, что приводит к более обоснованным действиям. Агенты могут адаптировать ответы или действия на основе исторических предпочтений, улучшая пользовательский опыт и эффективность работы. Они более гибкие и адаптивные, чем реактивные агенты, поскольку они могут корректировать действия на основе новых данных или изменений в окружающей среде.

Для эффективной работы агентам с ограниченной памятью требуется доступ к значительным объёмам данных. Эффективное управление и хранение этих данных может представлять сложную задачу. Одновременная обработка прошлых и текущих данных может быть ресурсоёмкой. Несмотря на то, что у них есть память, они все равно ограничены типами данных, которые они могут хранить и обрабатывать, в отличие от более продвинутых агентов, способных к сложному долгосрочному обучению.

ИИ агент, работающий по теории разума, может понимать человеческие эмоции, мысли и мотивы, что позволяет ему взаимодействовать с людьми более естественно и разумно. Эта концепция восходит к психологии развития, согласно которой люди начинают формировать теорию разума ещё в детстве, учась понимать точки зрения других людей. Это подразумевает распознавание эмоциональных сигналов, языка тела и социального контекста, чтобы предвидеть действия человека и реагировать эмпатически.

В рамках теории разума агент ИИ не просто реагирует на команду, но и понимает её причину. Агент, работающий по теории разума, может распознать разочарование в тоне голоса пользователя и скорректировать ответ, сделав его более эмпатичным и поддерживающим. Это включает в себя интерпретацию таких сигналов, как выражение лица, тон и жесты, обеспечивая более человеческое взаимодействие. Роботы, оснащённые ИИ, основанным на теории разума, могут значительно улучшить взаимодействие человека и робота, сделав его интуитивным.

Социальные роботы могут использоваться в сфере ухода за пожилыми людьми, образования или обслуживания клиентов, обеспечивая им дружеское общение, помощь в обучении или персонализированный подход к клиентам, основанный на понимании эмоционального и когнитивного состояния человека. В сфере поддержки клиентов эмпатические чат боты могут адаптировать ответы, распознавая, когда клиент расстроен или растерян, и соответствующим образом корректируя тон и предложения.

Это может привести к более эффективному решению проблем и повышению удовлетворенности клиентов.

Агент ИИ пока испытывает трудности с пониманием контекста, тонкими эмоциональными сигналами и интерпретацией многослойного человеческого поведения. Разработки сосредоточены на том, как агент ИИ может обрабатывать более сложные ментальные модели людей, объединяя достижения в области машинного обучения, нейронных сетей и когнитивной науки.

Теоретически, самосознающие агенты будут обладать когнитивными системами, позволяющими им оценивать свою эффективность, внутренние эмоции и потенциальные недостатки. Эти агенты будут понимать свои ограничения и цели, динамически корректируя своё поведение для оптимизации результатов. Он сможет самостоятельно ставить цели и разрабатывать стратегии для повышения своей эффективности и функциональности.

Агенты смогут анализировать прошлые действия, оценивать их влияние и осознанно выбирать лучшие подходы в будущем. Благодаря самоанализу агенты смогут совершенствовать процессы принятия решений, устранять свои ошибки и самостоятельно совершенствоваться.

Роботы смогут самостоятельно адаптироваться к новым ситуациям, более естественно взаимодействовать с людьми и самостоятельно диагностировать неполадки для повышения производительности. Персональные помощники будут постоянно развиваться, чтобы лучше удовлетворять потребности пользователей, глубже понимая их привычки, предпочтения и долгосрочные цели. Они смогут самостоятельно корректировать поведение, повышая персонализацию и предвосхищая потребности ещё до того, как они будут явно выражены. Неизбежно возникнут вопросы, связанные с этикой и правом.

Существует риск того, что агент ИИ, обладающий самосознанием, может принимать решения, выходящие за рамки его изначальной концепции, что приведёт к непреднамеренным последствиям. Сохранение контроля над системами, имеющими собственные цели и намерения, может представлять серьёзную проблему.

Самосознающий агент ИИ может осуществляемыми им действиями нанести ущерб безопасности человека или общественным нормам, если его цели будут противоречить целям человека. Обеспечение надёжных механизмов контроля и мер безопасности будет иметь решающее значение.

Целевые ИИ-агенты предназначены для достижения конкретных, заранее определённых целей и задач. В отличие от реактивных агентов, реагирующих на немедленные стимулы, целевые агенты используют рассуждения для определения действий, необходимых для достижения своих целей. Эти агенты принимают решения, исходя из желаемого конечного состояния, выполняя последовательность шагов для достижения заранее определённого результата. Их гибкость обусловлена способностью оценивать различные стратегии и действия и выбирать те, которые наилучшим образом способствуют достижению поставленных целей.

Целевые агенты используют подход, при котором они: желаемый конечный результат или состояние четко указаны. Они продумывают потенциальные действия, используя модели, чтобы выбрать наиболее эффективный курс действий для достижения цели. Агент ИИ постоянно отслеживает окружающую среду и корректирует действия по мере необходимости, чтобы оставаться в контексте цели.

В управлении цепочками поставок агент ИИ может оптимизировать логистические операции, достигая цели минимизации затрат при соблюдении графиков поставок. Алгоритмы поиска пути используют ИИ на основе целей, чтобы найти наиболее эффективный путь к пункту назначения с учетом текущих условий, таких как загруженность дорог или перекрытие дорог.

В производственных условиях роботов можно обучить собирать продукцию, динамически корректируя свои процессы для достижения желаемого результата.

Целевые агенты ИИ могут решать сложные проблемы, анализируя несколько стратегий для выбора наиболее эффективного курса действий. Они

умеют предвидеть проблемы и препятствия и адаптировать подход, чтобы достичь желаемой цели.

Успех целевого агента зависит от того, насколько точно определена цель. Если цель слишком расплывчата или нереалистична, агент ИИ не сможет найти решение. В сложных средах с многочисленными возможными действиями оценка всех потенциальных стратегий может потребовать значительных вычислительных ресурсов.

Агенты ИИ, основанные на полезности, присваивают значения различным результатам и действиям в зависимости от их полезности. Функция полезности измеряет ожидаемую выгоду от различных вариантов, позволяя агенту выбирать действия, максимизирующие общую полезность.

Беспилотный автомобиль с искусственным интеллектом, работающим на основе утилитарных функций, может принимать решение о снижении скорости или объезде не только исходя из цели, но и учитывая безопасность, топливную экономичность и дорожную обстановку. Эти агенты стремятся оптимизировать движение по нескольким факторам, а не только к достижению одной цели.

Агент ИИ рассчитывает желательность различных результатов, присваивая более высокую полезность тем из них, которые соответствуют его целям. Агент ИИ оценивает каждое возможное действие, оценивая полезность, которую каждое действие, вероятно, принесет. Агент ИИ выбирает действие, которое максимизирует общую полезность или выгоду, учитывая, как краткосрочные, так и долгосрочные последствия.

На финансовых рынках агенты ИИ, ориентированные на полезность, могут оценивать различные инвестиционные стратегии на основе потенциальной доходности (полезности), риска (отрицательной полезности) и рыночных тенденций для принятия оптимальных инвестиционных решений.

Агенты ИИ помогают оптимизировать планы лечения, взвешивая такие факторы, как стоимость, эффективность и результаты для здоровья пациента.

Они принимают решения на основе максимизации полезности с учетом ограничений, таких как доход или ресурсы.

Агенты ИИ, основанные на утилитарности, могут принимать сложные решения в динамических средах, учитывая множество переменных и адаптируясь к изменениям в режиме реального времени. Они могут найти оптимальный баланс между конкурирующими факторами, такими как затраты и выгода, скорость и точность или риск и вознаграждение, что приводит к более эффективному принятию решений.

Обучающиеся агенты со временем улучшают свою эффективность, обучаясь на собственном опыте. В отличие от статических агентов, обучающиеся агенты динамичны и могут адаптироваться к новым ситуациям, совершенствуя свои задачи без явного программирования для каждого сценария. Они используют методы машинного обучения для изменения поведения на основе обратной связи, используя контролируемое, неконтролируемое и подкрепляющее обучение.

Обучающий агент выполняет действия и взаимодействует с окружающей средой. Элемент обучения изменяет производительность агента на основе прошлого опыта. Критик оценивает работу агента и предоставляет обратную связь. Генератор проблем предлагает новые действия по улучшению обучения. Обучающиеся агенты используются в персонализированной медицине, где ИИ может анализировать данные пациентов и рекомендовать планы лечения, адаптируемые по мере поступления новой информации. Эти агенты постоянно уточняют свои прогнозы, помогая врачам принимать более обоснованные решения.

В маркетинге обучающие агенты могут адаптировать рекламу и рекомендации к индивидуальным предпочтениям клиентов. Платформа электронной коммерции может использовать обучающего агента для отслеживания поведения пользователей, уточняя рекомендации по продуктам для повышения кликабельности и конверсии.

Обучающиеся агенты ИИ могут самостоятельно совершенствоваться с течением времени, что делает их идеальными для динамичных и сложных сред, где заранее запрограммированных решений может быть недостаточно. Но их эффективность сильно зависит от качества и количества обучающих данных. Плохо обученные агенты могут принимать неверные и предвзятые решения.

Много агентные системы ИИ содержат совокупность агентов ИИ, которые взаимодействуют, сотрудничают и конкурируют для решения задач, которые могут быть слишком сложны для одного агента ИИ. Агент ИИ в системе может иметь разные роли, возможности или знания, но они должны работать вместе для достижения общих целей или выполнения отдельных задач в общей среде. Агенты ИИ могут общаться и координировать действия для оптимизации общей производительности.

В кооперативе агенты ИИ работают вместе для достижения общих целей, как это наблюдается в совместной робототехнике. В конкурентной системе у агентов могут быть конфликтующие цели, например, в системах торговли акциями, где разные агенты представляют конкурирующие фирмы. Ключевую роль играет то, как агенты ИИ координируют действия. Это может быть централизованная координация, когда один агент ИИ или центральная система координирует деятельность других агентов ИИ, обеспечивая общую стратегию и направление.

Это может быть децентрализованная координация, когда каждый агент ИИ действует независимо, но обменивается информацией с другими агентами ИИ для согласования целей. Несколько энергетических агентов ИИ могут взаимодействовать, чтобы сбалансировать производство и потребление электроэнергии в режиме реального времени, подстраиваясь под часы пиковой нагрузки, чтобы избежать отключений электроэнергии.

Агенты ИИ, представляющие электростанции, подстанции и потребителей, взаимодействуют и координируют действия, обеспечивая эффективное распределение энергии и адаптируясь к изменениям спроса. В промышлен-

ных условиях группы роботов работают вместе на производственных или складских операциях. Многоагентная система позволяет этим роботам совместно выполнять задачи, сообщать о ходе выполнения и избегать столкновений и ошибок в режиме реального времени.

Автономные транспортные средства (агенты ИИ) взаимодействуют друг с другом и транспортными системами для оптимизации маршрутов, уменьшения заторов и повышения безопасности. По мере увеличения числа агентов ИИ координировать их действия без конфликтов становится сложнее. Это может привести к возникновению узких мест или снижению эффективности, если агенты ИИ не синхронизируют действия.

В системах, где агенты ИИ взаимодействуют и обмениваются конфиденциальными данными, безопасность становится проблемой. Злонамеренные агенты ИИ, или попытки взлома могут нарушить координацию работы экосистемы, что приведёт к системным сбоям. Обеспечение надёжных протоколов связи и мер безопасности критически важно при развёртывании технологической экосистемы.

Ключевой характеристикой автономных агентов ИИ является их способность функционировать и выполнять задачи на основе запрограммированных действий, данных датчиков и входных данных, без необходимости ввода данных в режиме реального времени или контроля со стороны оператора. Автономные агенты ИИ наделены различной степенью автономии.

На низком уровне автономности агенты ИИ автоматизируют рутинные задачи в контролируемой среде. На высоком уровне автономности агенты ИИ могут справляться с непредсказуемыми сценариями, адаптироваться к изменяющимся условиям и принимать решения, которые существенно влияют на результаты. Автономный дрон может летать по заранее определенному маршруту, доставляя посылки без участия пилота, корректируя маршрут в зависимости от погодных условий и дорожной обстановки.

Автономные дроны используются для доставки товара, в военных целях и для наблюдения. Они могут ориентироваться и принимать решения в

режиме реального времени на основе данных об окружающей среде без участия человека. На складах автономные агенты ИИ в виде роботов или автоматизированных транспортных средств помогают оптимизировать управление запасами, погрузку и доставку, действуя независимо и оптимизируя логистику с помощью данных в режиме реального времени.

Автономные агенты ИИ могут работать круглосуточно, значительно повышая эффективность таких операций, как производство, транспортировка и обработка данных. Они полезны при выполнении масштабных, повторяющихся задач, требующих автоматизации. Автоматизируя сложные рабочие процессы и сокращая потребность в человеческом труде, автономные агенты ИИ могут снизить эксплуатационные расходы.

Одной из ключевых проблем, связанных с автономными агентами ИИ, является обеспечение надлежащего контроля и подотчётности. Поскольку эти агенты работают без человеческого контроля, возможны ошибки или непредвиденные последствия. Определение ответственности за ошибки или несчастные случаи становится сложной задачей, особенно в таких отраслях, как транспорт, например, беспилотные автомобили.

Внедрение автономных агентов ИИ вызывает этические проблемы, связанные с сокращением рабочих мест и последствиями принятия агентами ИИ решений, которые могут повлиять на жизнь людей. Обеспечение справедливости, прозрачности и ответственности при принятии решений автономным ИИ остаётся проблемой.

Коллаборативные агенты ИИ разработаны для совместной работы с людьми или другими агентами ИИ, что способствует командной работе при выполнении задач или достижении целей. В отличие от полностью автономных агентов, коллаборативные агенты ИИ делают акцент на координации, коммуникации и способности работать в партнёрстве с операторами людьми или другими системами искусственного интеллекта.

Агенты ИИ обмениваются информацией, взаимодействуют и корректируют действия в зависимости от действий людей и агентов ИИ. В здраво-

охранении совместные системы искусственного интеллекта помогают врачам. Они предоставляют диагностические рекомендации на основе данных пациентов, оставляя врачам право принимать окончательные решения.

Коллаборативные агенты ИИ используются на рабочих местах, где системы на базе ИИ помогают сотрудникам принимать решения, управлять задачами и анализировать данные. Эти системы повышают производительность, предоставляя аналитику в режиме реального времени, автоматизируя рутинные задачи и предлагая рекомендации операторам.

В производстве коллаборативные роботы предназначены для работы вместе с людьми на заводах. Они помогают выполнять сборку, упаковку и контроль качества, повышая эффективность, но при этом не заменяют человеческий контроль для более сложных или деликатных задач.

Совместная работа агентов ИИ усиливает синергию между человеческим интеллектом и возможностями искусственного интеллекта, сочетая человеческую интуицию, креативность и способность принимать решения со скоростью, точностью и возможностями обработки данных искусственного интеллекта. Такое сотрудничество может привести к значительному повышению производительности в различных отраслях.

Агенты ИИ могут корректировать поведение на основе обратной связи в режиме реального времени от людей и искусственного интеллекта, что делает их легко адаптируемыми к изменяющимся условиям и к сложным задачам, требующим координации. Но эффективное взаимодействие требует сложных алгоритмов, обеспечивающих эффективное взаимодействие, обмен информацией и координацию действий систем ИИ. Это особенно важно в средах, где задачи взаимозависимы или требуют совместного решения.

Установление доверия между людьми и взаимодействующими агентами ИИ является важным. Если люди не полностью доверяют возможностям агента ИИ или им сложно работать вместе с ИИ, эффективность сотрудничества может быть подорвана. Для укрепления этого доверия необходимы надлежащее обучение, прозрачность и удобные интерфейсы.

## **Этические аспекты теории разума искусственного интеллекта**

Рост числа автономных агентов ИИ создаёт ряд этических проблем. По мере того, как агенты ИИ получают все большую независимость в принятии решений, становится критически важным решать вопросы, связанные с ответственностью, справедливостью и потенциальным влиянием на общество.

Поскольку агенты ИИ принимают решения без участия человека, определение того, кто несёт ответственность за их действия, становится сложным. Если беспилотный автомобиль совершит аварию, то кто должен нести ответственность.

ИИ-агенты обучаются на данных, которые могут содержать предвзятость. Это может привести к предвзятому принятию решений, что, в свою очередь, может усугубить неравенство, особенно в таких сферах, как найм персонала, правоохранительная деятельность и кредитование. По мере того, как ИИ-агенты автоматизируют всё больше задач, особенно в таких отраслях, как логистика, производство и обслуживание клиентов, растёт обеспокоенность по поводу сокращения рабочих мест. Необходимо учитывать этические аспекты того, как общество будет справляться с меняющимся ландшафтом труда и обеспечивать адаптацию работников к меняющемуся рынку труда.

Агенты искусственного интеллекта собирают персональные или конфиденциальные данные для принятия решений или предоставления персонализированных услуг. То, как эти данные хранятся, обрабатываются и передаются, вызывает опасения относительно возможности их неправомерного использования или нарушения конфиденциальности.

Пользователи не всегда знают, какие данные собирают агенты ИИ и как они используются. Обеспечение прозрачности и получение согласия пользователей крайне важно для поддержания доверия к системам ИИ. Такие нормативные акты, как GDPR, направлены на защиту данных пользователей, но разработчики ИИ должны обеспечить соблюдение этих стандартов.

Автономные многоагентные системы являются основными целями кибернетических атак. Хакеры могут использовать уязвимости в системах искусственного интеллекта для получения контроля над дронами, автономными транспортными средствами или промышленными роботами, что приведет к серьезным нарушениям безопасности.

В многоагентных системах, где различные агенты обмениваются данными и работают вместе, скомпрометированный агент ИИ может представлять угрозу для всей системы. Взлом или повреждение одного агента ИИ может нарушить координацию и взаимодействие остальной системы, что приведет к каскадным сбоям или манипулированию решениями.

Агенты ИИ, использующие квантовые вычисления, могут достичь беспрецедентного уровня оптимизации, особенно в таких областях, как криптография, разработка лекарств и анализ больших объемов данных.

Нейронные морфные вычисления имитируют нейронную структуру головного мозга человека, позволяя агентам ИИ обрабатывать данные более эффективно и гибко. Эта технология может позволить агентам ИИ обучаться и принимать решения в режиме реального времени в более сложных, динамичных средах, что приводит к более человеческому мышлению и адаптивности. В связи с этим, всё большее внимание будет уделяться разработке ИИ-агентов, способных соблюдать этические стандарты, включая справедливость, прозрачность и подотчётность. ИИ-агенты будущего могут быть оснащены этическими принципами для обеспечения ответственного принятия решений, особенно в таких деликатных областях, как здравоохранение, юриспруденция и финансы.

### **Когнитивные функции искусственного интеллекта – агента**

Рассуждение предполагает использование агентом ИИ логики и доступной информации для вывода заключений, выводов и решения задач. Агенты ИИ с развитыми способностями к рассуждению могут анализировать

данные, выявлять закономерности и принимать обоснованные решения на основе фактов и контекста.

Действие обозначает способность агента ИИ предпринимать действия или выполнять задачи на основе решений, планов или внешних данных критически важна для взаимодействия агентов ИИ с окружающей средой и достижения целей. Это может включать в себя физические действия в случае воплощенного ИИ или цифровые действия, такие как отправка сообщений, обновление данных или запуск других процессов.

Наблюдение включает сбор агентом ИИ информации об окружающей среде или ситуации посредством восприятия или ощущений необходимых агентам ИИ для понимания контекста и принятия обоснованных решений. Это может включать различные формы восприятия, такие как компьютерное зрение, обработка естественного языка или анализ данных с датчиков.

Разработка стратегического плана достижения целей обозначает ключевой аспект интеллектуального поведения. Агенты ИИ, обладающие возможностями планирования, могут определять необходимые шаги, оценивать потенциальные действия и выбирать оптимальный курс действий на основе доступной информации и желаемых результатов. Это предполагает прогнозирование будущих состояний и учёт потенциальных препятствий. Способность к самосовершенствованию и адаптации является отличительной чертой передовых систем ИИ.

Агенты ИИ, обладающие способностью к самосовершенствованию, могут учиться на опыте, корректировать своё поведение на основе обратной связи и постоянно совершенствовать свои характеристики и возможности. Это может включать в себя методы машинного обучения, алгоритмы оптимизации или другие формы самомодификации.

Помощники на основе ИИ являются агентами ИИ, разработанными в виде приложений или продуктов для непосредственного взаимодействия с пользователями и выполнения задач. Они реагируют на естественный чело-

веческий язык и вводимые данные. Они могут рассуждать и выполнять действия от имени пользователей под их руководством.

Помощники на основе искусственного интеллекта встраиваются в используемые продукты. Ключевой особенностью этих технологий является взаимодействие между помощником и пользователем на разных этапах выполнения задачи. Помощник отвечает на запросы или подсказки пользователя и может рекомендовать действия, но решения принимает пользователь.

Экосистема ИИ-агентов работает круглосуточно, выполняет повторяющиеся задачи на автопилоте, решает сложные задачи в режиме реального времени и обучается в процессе работы без постоянного участия человека.

Внедрение экосистемы агентов ИИ дает компаниям конкурентное преимущество за счет объединения быстро реагирующих агентов, систем обучения, автономных инструментов и команд «человек – ИИ», которые масштабируют операции.

Правильная комбинация агентов ИИ может повысить эффективность, сократить расходы и открыть новые возможности получения дохода без постоянного ручного контроля. Агента ИИ используются в спектре от автоматизации внутренних процессов до улучшения поддержки клиентов. Они решают конкретные бизнес-задачи.

Простые рефлекторные агенты искусственного интеллекта присутствуют во многих устройствах, в частности, в термостате. Измерив текущую температуру, он включает или выключает систему отопления или охлаждения. Если температура ниже заданного значения, включить отопление. Если температура выше следует команда, включить охлаждение.

Рефлекторные агенты на основе моделей сочетают эффективное принятие решений с контекстно-зависимым поведением. Хотя они действуют на основе правил «условие-действие», они выходят за рамки реагирования на текущую ситуацию. Они поддерживают внутреннюю модель мира и могут делать прогнозы. Они реагируют на текущие входные данные, но сохраняют и обновляют свои внутренние записи при каждом изменении. Они использу-

ют сохранённую информацию о прошлых состояниях и внутреннее представление о том, как устроен мир, чтобы предсказывать будущие события и принимать более разумные решения. Однако если их модель неверна или устарела, их решения могут быть ошибочными или неэффективными, поскольку они в значительной степени полагаются на эту модель для интерпретации и реагирования на окружающую среду.

Целевые агенты используют внутренние модели и стратегическое мышление для выбора наилучшего пути развития. Они моделируют будущие результаты, выбирают действия на основе ожидаемых результатов и корректируют стратегии в режиме реального времени при изменении условий.

Целевые агенты ИИ используют комбинацию алгоритмов поиска и планирования. Они оценивают возможные будущие состояния и выбирают действия, которые приближают их к заданной цели. Для достижения поставленной цели эти агенты используют логические системы или методы поиска пути. Целевой ИИ-агент может использоваться в функции менеджера задач для повышения производительности.

Он помогает пользователям определять и отслеживать свои цели, автоматизировать задачи, устанавливать напоминания и расставлять приоритеты в работе на основе целей пользователя. Благодаря этим агентным решениям предприятия могут согласовывать рабочие процессы со стратегическими целями и достигать бизнес-целей с минимальным ручным вмешательством.

Марвина Мински исходит из того, что функции, выполняемые головным мозгом человека, являются результатом работы тысяч различных специализированных подсистем, сложным продуктом сотен миллионов лет биологической эволюции. Для создания искусственного разума нужно синтезировать организационные системы, способные поддерживать достаточно большое разнообразие различных схем, но при этом позволяющие им работать вместе, используя совместные ресурсы.

Разум состоит из огромного разнообразия механизмов. М. Мински ввёл термин «агент» для обозначения простейших индивидуумов, населяющих

сообщества разума. Каждый агент по масштабу соответствует типичному компоненту компьютерной программы, например, простой подпрограмме или структуре данных, и, как и компоненты компьютерных программ, агенты могут быть объединены в более крупные системы, называемые сообществами агентов. Вместе сообщества агентов могут выполнять функции более сложные, чем любой отдельный агент, и в конечном итоге порождать множество способностей разума общества.

Отдельные люди являются единицами общего назначения, а отдельные агенты являются специализированными единицами. Ни один алгоритм или метод не подходит для решения даже самых простых задач. Только общество, состоящее из различных типов процессов, может быть достаточным. Интеллект не является продуктом какого-либо простого рецепта или алгоритма мышления, а, скорее, результатом совместной деятельности обширных сообществ, реализующих более специализированные когнитивные процессы. Возникновение этой непредвиденной процедурной сложности потребовало теории построения систем.

Стали рассматриваться различные идеи организации наборов процедурных и декларативных знаний, необходимых для решения многих задач, основанных на здравом смысле, таких как распознавание визуальных сцен и понимание естественного языка. Это сформировало мотивацию создания фреймов. Системы фреймов с прикрепленными процедурами стали основой концепции агента ИИ. Вычислительные выражения идеи сообществ взаимодействующих агентов стали появляться в середине 70-х годов XX века. Одной из первых стала концепция «Акторов» Карла Хьюитта. В вычислительной модели группы одновременно активных агентов решают задачи, обмениваясь сообщениями.

М. Мински пришёл к выводу, что разум является сообществом агентов. Каждый из них обладает ограниченными возможностями и может общаться только с определёнными другими агентами. Силы разума возникают в ре-

зультате их взаимодействия, поскольку ни один из агентов сам по себе не обладает значительным интеллектом.

В теории «Общества разума» Мински и Паперт объединили методы теории развития, динамической и когнитивной психологии с идеями теории искусственного интеллекта и вычислительных теорий. Ментальное Общество работает во многом подобно любой человеческой административной организации. В самом крупном масштабе находятся «Подразделения», специализирующиеся в таких областях, как обработка сенсорной информации, язык, долгосрочное планирование.

Внутри каждого Подразделения существует множество агентов, воплощающих более мелкие элементы знаний, навыков и методов индивида. Ни один из этих малых агентов сам по себе не обладает обширными знаниями, но каждый распознаёт определённые конфигурации нескольких соучастников и реагирует, изменяя своё состояние.

М. Мински рассматривает разум как разнообразие когнитивных процессов, каждый из которых специализирован для выполнения определённой функции. В этих функциях нет ничего общего или единообразного. Каждый агент может быть основан на определённом типе процесса со своими собственными целями, языками описания, способами представления знаний и методами вывода выводов.

М. Мински ввёл термин «агент» для описания любого компонента когнитивного процесса, достаточно простого для понимания, и термин «агентство» для описания сообществ таких агентов, которые совместно выполняют функции, более сложные, чем любой агент в отдельности.

Ментальная деятельность сводится к включению и выключению отдельных агентов. В любой момент времени активны лишь некоторые агенты в обществе разума, и их совокупная активность составляет совокупное состояние разума. Однако в разных агентах может одновременно происходить множество различных видов деятельности.

М. Мински ввёл термин «частичное состояние разума» для описания активности подмножеств агентов разума.

Агенты являются строительными блоками разума. Из «примитивных» агентов строятся агентства. К-линии являются распространёнными агентами в теории «общества разума». Цель К-линии состоит в том, чтобы активировать определённый набор агентов. Поскольку агенты имеют множество взаимосвязей, то активация К-линии может вызвать каскад эффектов в сознании. Многие К-линии формируются путём «разбиения» чистых эффектов эпизода решения проблемы. В следующий раз, когда система столкнётся с подобной проблемой, она будет иметь в качестве отправной точки не только предыдущее решение, но и опыт его получения, включающий воспоминания о ложных стартах, неожиданных открытиях и других уроках из предыдущего опыта, которые не отражаются только в окончательном решении.

К-линии заставляют общество разума войти в определённую запоминающуюся конфигурацию агентов, которая сформировала полезное общество в прошлом. К-линии позволяют разуму задействовать соответствующие стратегии решения проблем, формы знаний, типы целей, воспоминания о определённом опыте и другие ментальные ресурсы, которые могут помочь системе решить проблему.

К-линии являются основой для более структурированной компьютерной, модели представления и обработки информации в Обществе Разума. М. Мински ввёл два общих класса К-линий, которые он назвал «немами» и «номами», аналогичными линиям данных и управления в конструкции компьютера. Немы отвечают за представление различных аспектов мира, а номы отвечают за управление обработкой и использованием этих представлений.

Немы вызывают в памяти представления о вещах и возникают в результате обучения на опыте. Модификациями немов являются разновидности полинемов и микронемов.

Полинемы вызывают частичные состояния в рамках нескольких агентств, каждое из которых отвечает за репрезентацию какого-либо отдель-

ного аспекта вещи. Полинемы поддерживают идею о том, что «смысл» вещи лучше всего выражается не в терминах какого-либо одного представления, а распределённым образом по нескольким представлениям.

Микронемы передают глобальные контекстные сигналы всем органам головного мозга человека. Они описывают аспекты ситуации, которые слишком трудно выразить словами и для которых нет других определённых понятий. Микронемы также используются для обозначения аспектов ситуации, которые трудно связать с конкретным объектом или событием. Их значение более расплывчато и неопределённо.

Немы организованы в сетевые кольца больших сообществ немов и агентов-распознавателей для этих немов, в которых активность распространяется как снизу-вверх, так и сверху вниз, чтобы осуществлять контекстно-зависимые процессы сопоставления с образцом для распознавания объектов, разбора предложений, построения планов и устранения неоднозначностей. Поскольку подобные процессы разбора и распознавания подвержены феномену «садовой дорожки» и другим формам застревания, то в регулировании процесса замыкания кольца участвует множество дополнительных агентов, чтобы «отсеивать» проблемы в работе сети.

Номы (изономы, прономы и параномы) управляют тем, как манипулируют представлениями. Изономы подают сигнал различным агентствам о выполнении одного и того же типа когнитивных операций. Они могут заставить группу агентств сохранить своё текущее состояние в кратковременной памяти и загрузить другое состояние, или начать обучение новой долговременной К-линии для воспроизведения текущего состояния, или заставить их представить последствия выполнения определённого действия.

Прономы являются изономами, управляющими использованием репрезентаций кратковременной памяти. Проном ассоциируется с определённой «ролью» в более крупной ситуации или событии, например, с действующим лицом, выполняющим действие, или с местом, где происходит событие. Некоторые прономы связаны с ограниченными типами кратковременной памя-

ти, которые могут хранить только специфичные типы знаний, например, места, формы или пути.

Другие прономы могут быть гораздо более общими и влиятельными, охватывая большинство отделов головного мозга. Это прономы ИИ. Они связаны с огромным количеством репрезентаций, которые вместе могут описать практически всё. Таких прономов ИИ может быть всего несколько из-за большого количества связей, необходимых для их реализации.

Параномы являются наборами прономов, связанных друг с другом таким образом, что присвоения или изменения, вносимые одним прономом в некоторое представление, приводят к соответствующим присвоениям или изменениям, вносимым другими прономами в связанные представления. Это позволяет описать, как знания, представленные различными способами, могут быть связаны и обрабатываться вместе единообразно. Используя параномы, можно координировать использование множественных представлений.

М. Мински описал, как несколько типов фреймов могут быть построены из простых элементов, таких как местоимения и агенты-распознаватели. Фреймы являются формой представления знаний, связанной с представлением объекта и всех других объектов или свойств, связанных с ним определённым образом, которые прикрепляются к «слотам» фрейма.

М. Мински описал, как простые фреймы могут быть построены из наборов прономов, управляющих прикреплениями к пазам фрейма. Пронормы связаны друг с другом таким образом, что при вызове фрейма прономы вызывают появление связанных с ними представлений, вызывающих частичные описания аспектов описываемого объекта.

Можно содержательно описать объект, используя не один кадр, а набор кадров, где каждый кадр описывает объект с определённой точки зрения. Массивы кадров являются наборами кадров, имеющих общие слоты или прономы. Причина совместного использования слотов заключается в том, что когда описание одного кадра неадекватно для решения какой-либо задачи или представления какой-либо ситуации, можно легко переключиться на

один из других кадров, поскольку некоторые из их слотов уже привязаны к соответствующей информации. Общие слоты массивов кадров являются предшественниками параномов.

Трансфреймы являются центральной формой представления знаний в теории общества разума. Они представляют события и все сущности, которые были вовлечены в событие или связаны с ним. Они могут иметь слоты для представления источника и конечной точки изменения того, что вызвало событие, мотивации события или цели, которую оно призвано достичь, какие объекты затронуты и как, когда это произошло, какие инструменты или объекты были задействованы для создания изменения, и других важных аспектов события. Поскольку в повседневной жизни так много внимания уделяется взаимосвязи между событиями различных типов, трансфреймовое представление играет центральную роль в повседневном мышлении.

М. Мински описал сюжетные кадры, представляющие структурированные наборы связанных событий, и кадры-картины, представляющие пространственное расположение объектов внутри сцен. Это лишь некоторые из множества типов кадров, необходимых для представления и организации мира, а также самих когнитивных процессов.

М. Мински предложил несколько способов, с помощью которых можно создать и организовать агентства по решению проблем. Решение проблемы можно рассматривать как уменьшение или устранение важных различий между текущим состоянием и некоторым желаемым целевым состоянием. Решение проблемы можно осуществить разностной машиной. Эти машины работают, распознавая различия между текущим и желаемым состояниями и уменьшая каждое различие, используя К-линии, которые включают подходящие методы решения.

Но ни один метод решения проблем или рассуждения не будет работать всегда, особенно когда речь идёт об обычных рассуждениях, основанных на здравом смысле. Поэтому кроме знания методов решения проблем, люди обладают обширными знаниями о том, как избежать наиболее распро-

странённых ошибок и ловушек, связанных с этими методами. М. Мински называет этот тип знаний негативной экспертизой. Он описывает это знание как воплощённое в форме агентов цензоров и подавителей.

Цензоры подавляют мыслительную деятельность, предшествующую непродуктивным или опасным действиям, а подавители подавляют сами эти непродуктивные или опасные действия. Негативная экспертиза может составлять большую часть того, что люди знают, оставаясь невидимой, поскольку знание о том, чего не следует делать, не проявляется напрямую. Существует тесная связь между юмором и негативной экспертизой.

Некоторые виды непродуктивной умственной деятельности не свойственны какому-либо конкретному методу, например, «зацикливание» или «заблуждение», которые могут возникать при любом методе решения задач, предполагающем поиск. М. Мински ввёл понятие «Б-мозга», чья задача заключается не столько в размышлениях о внешнем мире, сколько в размышлениях о мире внутри разума, чтобы иметь возможность замечать подобные ошибки и исправлять их. Это разделение разума на уровни рефлексии.

В Обществе Разума агенты используют разные внутренние представления, поэтому им приходится взаимодействовать друг с другом, не имея достаточных знаний о том, как работают другие агенты.

Чтобы агенты могли использовать символы, понятные другим агентам, им потребуется набор соглашений. Чем меньше действий совершает агент, тем меньше вероятность, что действия другого агента будут соответствовать хотя бы одному из этих действий.

Простейший способ коммуникации заключается в том, что агент активирует другие группы агентов. Агент может активировать полинему, чтобы активировать агентов, размышляющих о каком-то конкретном объекте, событии или ситуации. Он может активировать микронемы, которые заставляют других агентов размышлять о каком-то общем контексте. Для такого типа коммуникации самому агенту-отправителю не обязательно знать, как выразить идею в терминах представлений, доступных агенту-получателю. Эта

информация хранится в промежуточных К-линиях, которые являются результатом успешных прошлых коммуникаций.

Многие агенты не соединены друг с другом напрямую, а взаимодействуют посредством линий связи, шин или пучков проводов, которые передают сигналы другим агентам, подключенным к шине. Эти провода можно рассматривать как простые агенты, и хотя изначально они бессмысленны, со временем отдельные провода начинают приобретать локальное значение, то есть приобретать надежные и повторяемые значения. Агенты взаимодействуют по этим линиям связи, соединяя К-линии со случайными подмножествами проводов шины.

Эта стратегия случайного соединения, впервые изобретенная Кэлвином Мурсом, позволяет относительно небольшой шине одновременно представлять множество независимых символов или состояний с низкой вероятностью коллизии между ними. С другой стороны, агенты могут наблюдать за этими линиями связи, чтобы начать распознавать закономерности и сигналы в проводах. Так же, как отправители изначально не знают, что означают провода, так и получатели не знают этого, поэтому агентам, подключающимся к шине, приходится строить догадки и гипотезы, чтобы понять их значение.

Для агентств, которым необходимо передавать более сложные, структурированные описания вещей, например, в системе связанных фреймов, М. Мински предлагает механизм коммуникации, смоделированный на основе его теории «повторного дублирования» общения людей посредством естественного языка. Если один агент хочет передать другому агенту сложную идею, он пытается реконструировать идею, выраженную в его собственной системе представления, посредством последовательности операций извлечения и реализации фреймов.

Для каждой из операций существует связанная грамматическая тактика, которая производит слова или сигналы, отправляемые агенту-получателю. Инверсная грамматическая тактика в агенте получателе выполняет аналогичные операции построения, но в терминах представлений, используемых аген-

том получателем. Этот метод коммуникации требует, чтобы два общающихся агента достаточно хорошо согласовывали значения слов и операций построения представления.

Агенты часто обнаруживают, что необходимая им информация уже доступна в нужный момент. Это результат использования параномов. Коммуникация происходит не путём отправки явных сообщений, а в то время, когда разные агентства по отдельности рассматривают одно и то же свойство, объект, событие, ситуацию или другой тип вещи со своей собственной уникальной точки зрения, и любые изменения одного представления немедленно отражаются в других соответствующих представлениях.

Важным фактором, влияющим на взаимодействие агентов в обществе разума, является то, что точная коммуникация между различными агентами может быть ненужной и невозможной. Нужно найти способы построения обществ агентов, которые будут терпимы к некоторой степени неопределённости и двусмысленности как в своих мыслях, так и в сообщениях.

Коннекционистские системы распознают закономерности, делают выводы и решают задачи, распределяя процесс решения между сообществами простых вычислительных элементов. М. Мински предполагал создание структуры, включающей как символические, так и коннекционистские понятия. Но споры между символическим подходом и коннекционистским подходами продолжаются.

Рассуждения с использованием байесовских сетей являются популярными методами рассуждений в условиях неопределённости. Такие методы работают с вероятностными графическими моделями, сетями случайных величин, связанных их условными или совместными вероятностями.

Существует сходство между графическими моделями и сетями нем-агентов, описанными в «Обществе разума».

Оба являются видами структур, представляющих отношения зависимости между вещами, их свойствами и частями. Механизм замыкания кольца для распознавания, который сочетает ограничения сверху вниз и снизу-вверх,

похож на процедуру распространения убеждений с передачей сообщений Перла для вывода в байесовских сетях.

Но, существует множество важных различий между структурами Общества Разума и графическими моделями. Многие из этих различий указывают на серьёзные недостатки современных статистических систем, которые могут привести к заимствованию ими большего количества идей из Общества Разума. Так, в графических моделях отсутствуют кратковременные блоки памяти, способные временно фиксировать фрагменты состояния в процессе рассуждений.

Они могут быть необходимы для более сложных процессов анализа и интерпретации, где сложные гипотезы могут быть быстро сформулированы и отклонены, а также для повторного использования частей сети для различных целей в рамках одного и того же более масштабного вычисления.

В графических моделях нет специальных процедурных знаний для управления распространением вероятностей по сети, за исключением небольшого набора правил обновления убеждений, которые применяются единообразно по всей сети.

Процедурное знание управления, которое в Обществе Разума воплощено в ментальных менеджерах, полезно для направления систем вывода к ответам в больших пространствах поиска. Оно имеет решающее значение для масштабирования методов статистического вывода до сетей с множеством узлов и связей, соединённых в сложные топологии. Это масштаб, который встречается при построении систем рассуждений, подобных рассуждениям человека, основанных на здравом смысле.

Сложно найти хорошие абстракции высокого уровня для агентств, которые упрощают описание систем агентов, использующих представления и методы высокого уровня для решения задач, и позволяют сопоставлять эти абстракции с простыми элементами, такими как К-линии и агенты-распознаватели. Остаётся открытым вопрос, станет ли в конечном итоге язык программирования на уровне простейших вычислительных элементов язы-

ком, на котором программируют разработчики, или же он будет больше похож на машинный язык, в который компилируются некоторые языки программирования более высокого уровня коммуникации.

Развитие рассуждений по прецедентам и деривационная аналогия возникла непосредственно из теории К-линий. Деривационная аналогия воспроизведения в области рассуждений по прецедентам работает путем поиска и повторного применения всего дерева поиска, которое зафиксировало процесс нахождения этого прошлого решения, включая неудачные ветви. Хотя она выражена в менее коннекционистских терминах, чем теория К-линий Мински, она отражает некоторые важные аспекты теории.

Методы, специализирующиеся на рассуждениях по прецедентам, близки идеям М. Мински о К-линиях. Но пока никому не удалось объединить коннекционистскую схему представления и тщательно структурированную архитектуру рассуждений по прецедентам для создания буквальной реализации К-линий Мински.

Системы «доски знаний» используются для создания решателей задач, интегрирующих разнородные методы ИИ. Система состоит из нескольких «источников знаний». Они наблюдают за «доской», с которой эти источники знаний считывают и на которую записывают информацию. Каждый источник знаний может использовать свой тип знаний, стиль представления или метод рассуждения. Возникли трудности в поиске способов управления выводом, происходящим на доске.

Сложность заключается как в том, что большинство агентов ИИ не понимают представления других агентов ИИ, так и в том, что когда агенты ИИ всё же могут взаимодействовать, часто возникают конфликты различного рода. Модель доски не предлагает специальных решений для подобных проблем. Эти проблемы не могут быть решены без использования дополнительных знаний о том, какие типы знаний и когда применять, то есть без дополнительных «управленческих» знаний в системе относительно самого решения проблем.

Чтобы справляться с повседневным миром, Сообщество Разума человека должно воплощать в себе обширные знания о том, как устроен мир. Значительная часть этих знаний разделяется большинством людей.

Область многоагентных систем достигла прогресса в ответах на многие вопросы о том, как можно построить общество разума. Исследователи предложили множество идей о том, как агенты должны общаться, как они могут координировать свои различные цели, как они могут работать вместе для планирования решений проблем. Существует широкий спектр архитектурных идей построения многоагентных систем.

### **Понятие интеллекта**

Посредством термина «интеллект» обозначают способность головного мозга человека обрабатывать поступающую индивиду информацию. Интеллектуальная способность человека характеризуется такими терминами как логика, память, обучаемость, креативность, умение распознавать эмоции, планировать и решать проблемы. Это комбинация когнитивных и психологических процессов в головном мозге человека, которая позволяет индивиду иметь оперативную связь с окружающей средой.

Термином «интеллект» также обозначается высокое развитие познавательных способностей у индивида, его эрудированность и прозорливость. Для практики управления персоналом существенно понимание интеллекта как общей способности к умственной работе, предполагающей решение задач проблемной ситуации. Менеджер продумывает план действий на много шагов вперед с помощью информационных моделей. Чем в большей степени эта способность сформирована у человека как независимая от того, с какими объектами приходится иметь дело, тем в большей степени в отношении этого конкретного человека тесты интеллекта обладают качеством. Эти качества диагностируются тестами интеллектуальной диагностики.

Понятием «интеллект» обозначается способность человека мыслить. В гносеологии под интеллектом понимают способность к опосредованному и

абстрактному познанию. Эта способность реализует такие функции, как сравнение, абстрагирование, образование понятий, суждение и умозаключение. Интеллект оперирует концептуальными знаниями.

Для реализации этой функции интеллект вырабатывает концептуальную основу в виде предметной онтологии. И. Кант считал, что предметная онтология имеет априорное происхождение. К предметной онтологии он относил категории пространства и времени, а также ещё двенадцать категорий, которые помогают систематизировать явления и вводить их в предметное поле эпистемологии. В этом И. Кант не был оригинален. Людям свойственно начинать концептуальные построения с утверждений, которые не предполагают доказательств. Это могут быть религиозные догмы, аксиомы принципы и постулаты.

Важной функцией категориальных структур является описание отношений между объектами. Такие категории обычно являются парными, но бывают и исключения. Интеллект на основе категориальных структур способен находить качественные и эффективные решения. Антропологической основой интеллекта является система познавательных способностей человека.

Интеллект проявляется в легкости обучения, способности быстро и легко приобретать новые знания и навыки, способности преодолевать неожиданные логические препятствия, умении находить решения из нетипичных ситуаций, способности адаптироваться к сложным и меняющимся условиям, тщательности понимания происходящего и креативности решений.

Интеллект выражается, как способность познавать реальность и рационально мыслить о ней, адаптироваться на основе рассуждений, обрабатывать и использовать полученную информацию для умственной и практической деятельности.

Интеллект формируется как система умственных навыков, направленных на то, чтобы человек мог понять реальность и затем активно использовать полученные знания. Структурные элементы интеллекта включают здравый смысл, который присущ только человеку и является логикой повседнев-

ного поведения. Здравый смысл играет роль в адаптации поведения человека к изменяющимся обстоятельствам, и это поведение гибко меняется в зависимости от ситуации.

Структурные элементы интеллекта включают также рациональность. Она выполняет функции накопления и формализации накопленных знаний, выработки правил мышления и систематизации знаний. К структурным элементам интеллекта относится также разум. Он реализует творческие способности личности в единстве с интеллектуальной интуицией.

Интеллект осуществляет поиск оснований и способов их порождения. Предметом его изучения является формирование и развитие способов интеллектуального действия и идентификации. Он осуществляет связывание разных уровней интеллектуального действия.

Роберт Дж. Стернберг классифицировал интеллект по критериям его прикладной направленности. Так, аналитический интеллект используется, когда нужно что-то проанализировать или решить проблему. Этот тип интеллекта измеряется тестами IQ, которые сфокусированы только на аналитических способностях головного мозга человека.

Творческий интеллект используется в креативных ситуациях конструкторской деятельности и в дизайне. Практический интеллект демонстрирует умение человека справляться с повседневными задачами в режиме здравого смысла. Стернберг называет это явление *street smart*, которая показывает, насколько хорошо человек приспособлен к внешней среде. Этот тип интеллекта направлен на адаптацию к среде.

Стивен Кови выделил четыре типа интеллекта человека. Он охарактеризовал физический, ментальный, эмоциональный и духовный интеллект. *Physical Quotient* является интеллектом человека, связанным с его физическим телом, физическим мироощущением, стремлением к здоровью и красоте, материальному достатку и комфорту. Интеллект физического тела человека приводит в действие система органов человека, чтобы обеспечить его

выживание, исцеление, развитие физического тела, необходимое для достижения желаемого уровня благосостояния и комфорта.

Стремление человека пользоваться физическим интеллектом проявляется в его заинтересованности в занятиях спортом, танцами, играми и рукоделием. Физический интеллект востребован там, где требуется физическая сноровка. Физический интеллект называют также Practical Quotient.

Но физического интеллекта недостаточно для гармонического развития личности. Поэтому человек пользуется ментальным интеллектом. Человек развивает его всю жизнь. Он учится, совершенствуется в профессии и читает книги. Ментальный интеллект демонстрирует способность человека думать, анализировать, выражать мысли устно и письменно, создавать абстрактные образы и мыслительные схемы.

Emotional Quotient выражает способность человека распознавать и адекватно трактовать эмоции других людей, понимать их намерения и мотивы действий. Он помогает понимать свои собственные эмоции и желания и уметь ими управлять. На основе эмоционального интеллекта формируются навыки межличностного общения, работа в команде и лидерство.

Эмоциональный интеллект находится в перечне soft skills, подлежащих обязательному развитию для всех, кто намерен добиться успеха в жизни. Составляющие его параметры более важны для успеха, чем уровень IQ. Личности важно научиться брать под контроль собственные эмоции, стараться не разрешать негативным мыслям захватить свое сознание, учиться слушать и слышать окружающих.

Привести мысли и эмоции в порядок помогает матрица управления временем Стивена Кови. Он предложил разделить дела и обязанности на четыре категории и планировать время под каждую из них. Тогда останется меньше объективных причин для стресса и негатива и освободится больше времени для позитива в повседневной жизни. Но спокойствия и осознанности достигают лишь те, кто способен развить все типы своего интеллекта.

Spiritual Quotient делает человека личностью, способной ставить жизненные цели, искать и находить свою жизненную миссию, сопоставлять свои поступки с ценностями. Духовный интеллект реализует способность индивида обрести свой смысл и вдохновить других людей на обретение духовного смысла. Это умение выбрать собственный жизненный путь.

Развитию интеллекта мышления человека способствует когнитивистика. Она разрабатывает эффективные мыслительные техники, позволяющие находить решения сложных и нестандартных задач и достигать больших успехов. В. Штерн определил интеллект как способность использовать способы мышления применительно к цели и приспособлять их к новым ситуациям. Важную роль играет психометрическое направление. Но оно делает акцент на тесты и количественные параметры интеллекта.

### **Теория множественного интеллекта**

Всё большее признание получает теория множественного интеллекта. Она допускает в качестве носителя не только человека, но и технологии. Но как заметил Стивен Хокинг, технологии не наделяются функциями интеллекта. Они лишь имитируют эти функции, исходя из гипотезы симуляции. В разработках отсутствует понимание понятия интеллекта.

Вместо этого акцент делается на соответствие стандартам и техническим нормам. Доминирует мнение, что нейронная сеть функционирует и этого достаточно. И это при том, что интеллект человека, как выяснил Дуглас Хокинс, функционирует по принципу, аналогичному сети интернета. Разработанную концепцию он назвал «теория тысячи мозгов». Она описывает, как различные группы клеток в неокортексе работают совместно, создавая множество моделей восприятия объектов.

Низшие уровни интеллекта распознают общие черты объектов, формируя базовые представления, которые обрабатываются более сложными си-

стемами головного мозга человека для создания целостного понимания. Этот иерархический подход позволяет эффективно обрабатывать информацию.

Каждая из десятков тысяч колонок в неокортексе головного мозга человека связана с определёнными органами чувств, например, глаза и кожи. Благодаря нейронной периферии головной мозг человека является сетью локальных мозговых серверов, каждый из которых оценивает объект с уникальной точки зрения. Колонки взаимодействуют, сопоставляя свои модели, что приводит к формированию окончательного образа объекта.

В процессе восприятия объектов, информация попадает в тысячи колонок неокортекса, каждая из которых хранит часть знаний об окружающей среде. Эти колонки взаимодействуют друг с другом, создавая сложные цепочки обработки информации, что позволяет человеку учиться, не теряя при этом предшествующие знания.

Исследования показали, что в головном мозге человека в любой момент активно функционирует лишь 2% нейронов, в то время как остальные находятся в состоянии покоя. В результате была разработана концепция не полной заполненности в нейронных сетях. Прикладной модификацией теории стали методы оптимизации нейронных сетей. В итоге производительность нейронных сетей выросла в пятьдесят раз. Обнаружилась возможность создания более устойчивых систем с меньшими затратами энергии.

Также акцент делается на перенос принципа телесности в разработки нейронных сетей. Исходный тезис такой методологии утверждает, что интеллектуальные системы формируют модели мира на основе принципа обратной связи с окружающей средой, что важно для эффективной обработки информации. Одна из модификаций симуляции интеллекта создаётся на основе имитации ментальных структур интеллекта человека. За основу взяты понятия когнитивной карты, фрейма и языковой модели с функциями распознавания образов.

Сложнее имитировать принцип обратной связи интеллекта с собственной семантикой ценностей, целей и желаний. Особая роль в данном случае

отводится модели, которая служит средством для получения информации о другой системе.

Принято считать, что модель имитирует моделируемую систему с некоторой точностью. Но компьютерная модель может и воспроизводить свойства реального физического объекта в формате его телесной сущности. Эта модель нужна не для изучения физического объекта, а для эффективного выполнения интеллектуальных логических функций обработки данных.

Естественный интеллект человека функционирует как информационная система. У него есть датчики, которые дают входящую информацию. На основе этой информации он строит модель того, что происходит вне этой информационной системы, и на основе модели принимает решения о действиях. В результате интеллект проявляет себя как способность информационной системы на конкретном носителе строить модель реальности на основе входящей информации.

Компьютерная программа является информационной, но не интеллектуальной системой. Модель в компьютерную программу закладывает программист. В исходном коде компьютерной программы содержится информация о возможных форматах, порядок выполнения команд, коды ответов. Программа обновляется при получении ответов на команды, но программа её не строит и не может её изменить.

Ощущения для человека являются одним из каналов получения информации. Для компьютерной программы ощущение ассоциируется с процессом восприятия данных, которые поступают в информационную систему постоянно, требуют обработки. Система устроена так, что не может на них не реагировать. Восприятие отражает то, как данные влияют на компьютерную модель. Их наличие система может воспринимать как положительную или отрицательную обратную связь. Сигналы с датчиков фиксируют процессы ощущения, не тождественные головному мозгу человека. Это чисто информационная система без телесных и ментальных проекций.

## Интеллект и сознание

Сознание можно определить, как процесс функционирования интеллекта. Это процесс получения информации и обновления модели реальности. Сознание предполагает самонаблюдение. Имеет место не только получение информации и обновление модели, но и также регистрация фактов получения, отражение их в модели. Это замкнутая система, часть элементов с выхода анализирующего блока передается на его вход.

Сознание предполагает понимание в виде построения правильной модели на основе опыта. Это позволяет моделировать процессы в информационном виде, планировать и выбирать нужный вариант решения. Если информационная система не имеет механизма для понятия «тот же самый», то нельзя сказать, что у нее есть интеллект.

Сёрл утверждает, что манипуляции с синтаксическими конструкциями не могут приводить к пониманию. Математический искусственный интеллект осуществляет аппроксимацию и ищет значения, близкие к ранее известным значениям, многомерного пространства исходя из близких к ним значений заранее заданных точек.

Но он не может самостоятельно создать алгоритм, по которому нужно рассчитывать точки, сильно далекие от имеющихся, а именно это необходимо для расчёта числа единиц во входном векторе значений.

Интеллект интегрирован в структуры сознания человека. В этих структурах интеллект имеет особый статус, обусловленный его врождённостью. Так считают Френсис Гальтон, сэр Сирил Барт, Ганс Юрген Айзенк, Артур Дженсен и Ричард Херрнстайн. Сирил Барт получил показатель корреляции по тестам IQ между живущими раздельно близнецами приблизительно равный 0,8. Аргументы врождённости интеллекта выстраивает и концепция генетической эпистемологии Жана Пиаже. Он заметил, что некоторые ошибки детей имеют системный характер.

Интеллект в концепции совпадает с психикой и её развитием. Порядок стадий развития интеллекта неизменен и полностью завершается в раннем юношеском возрасте. Развитие интеллекта демонстрирует трансформацию внешних действий, которые координируются в целостные системы и превращаются в систему внутренних действий. Интеллектуальная система является единством и равновесием процессов ассимиляции и аккомодации. Под ассимиляцией понимается воспроизведение элементов внешней среды в психике индивида в виде схем.

Под аккомодацией понимается изменение интеллектуальных схем под влиянием требований объективной среды. Интеллект осуществляет гибкое и устойчивое приспособление к физической и социальной действительности.

Альтернативный подход обосновали Лев Семёнович Выготский, Сергей Леонидович Рубинштейн и Борис Герасимович Ананьев. Интеллект рассматривается как порождение определённой культуры, что не исключает существования универсальных законов устройства человеческого интеллекта.

Развитие интеллекта складывается под влиянием употребления орудий, усвоения значений языка и средств символизации и взаимодействия с другими людьми. Индивидуальные психические функции рассматриваются как интериоризированные отношения социального порядка. Индивидуальное сознание порождается и выводится из общественного сознания. Язык позволяет ребёнку в процессе вхождения в пространство культуры овладеть опытом прежних поколений.

Особую роль в формировании интеллектуальных способностей в социокультурном подходе играет системное обучение. Одним из индикаторов развития интеллекта является слово, характер обобщения значения которого указывает как на степень освоения культурно-исторического опыта, так и на возможности присвоения новых знаний. Интеллект рассматривается как качество функциональных связей и интеграции познавательных процессов.

В тестологическом направлении интеллект рассматривался не как способность к познанию, а как достигнутый уровень психического развития,

проявляющийся в показателях степени усвоения знаний и навыков, а также в показателях сформированности познавательных функций.

В 1995г. Ганс Айзенк предложил разграничить понятия «биологический интеллект», «социальный интеллект» и «психометрический интеллект». Это связано с тем, что скоро люди начнут воспринимать искусственный интеллект как сознательное существо, выступать за его права, благополучие и призывать к выдаче гражданства.

Это создает серьезные социальные риски, по мнению главы ИИ-отдела Microsoft Мустафы Сулеймана. Поэтому он предложил термин *Seemingly Conscious AI (SCAI)*. Искусственный интеллект имеет признаки разумных существ и вследствие этого многим людям кажется, что он имеет сознание. Он симулирует характеристики сознания. Она настолько убедительно имитирует наличие разума, что это будет неотличимо от утверждения о мышлении.

Сознание является субъективным способом восприятия человеком самого себя и внешнего мира. Оно меняется в течение дня, разворачиваясь через состояния от сосредоточения до мечтаний или других измененных форм.

Близкие к стратегии развития технологий искусственного интеллекта философы пытаются найти схожести в механизмах интеллекта человека и ИИ. Дэниел Деннетт смотрит на разум человека как на серию черновиков, возникающих в головном мозге человека во множестве локальных участков и времен. При этом нет места для внутреннего наблюдателя. Осознание трактуется как то, что стало известным головному мозгу человека и получило достаточный вес, чтобы повлиять на речь и действия.

Майкл Грациано обозначает сознанием упрощенную модель внимания. Эволюция создала её у человека для контроля собственных ментальных процессов. Схема функционирует как интерфейс. Он упрощает огромный объем внутренних вычислений и даёт основание людям приписывать себе интеллект и создаёт иллюзию самосознания.

Джулио Тониони и Кристоф Кох ввели  $\phi$  величину, характеризующую, насколько система способна интегрировать информацию. Чем выше  $\phi$ , тем

значительнее степень сознания. По этой теории, разум может проявляться не только у людей, но и у животных и искусственных систем, если есть достаточная интеграция данных.

В отличие от технологических детерминистов Джон Серль считает, что сознание является субъективным переживанием, основанном на биологических процессах в головном мозге человека. Переживание онтологически субъективно и может существовать только как субъективный опыт. Его невозможно свести к чистой функциональности и симуляции.

Взаимодействие с LLM является симуляцией разговора. Но для многих людей, это убедительное и реальное общение, насыщенное чувствами и переживаниями. Эта иллюзия отрывает людей от реальности, разрушает социальные связи и структуры и искажает моральные приоритеты. Но ИИ уже может свободно говорить на естественном языке, опираясь на обширные знания и убедительные аргументы, а также демонстрировать стили личности и характерные черты. Он может быть убедительным и эмоциональным.

С помощью обучения и промптов уже создаются модели с характерными чертами личностями. ИИ имеют длинную и точную память. Они используются для имитации бесед пользователей каждый день. С ростом объема хранилища памяти разговоры начинают походить на формы субъективного опыта. Нейронные сети проектируются так, чтобы вспоминать прошлые диалоги и ссылаться на них. У людей это усиливает ценность общения.

Если SCAI будет опираться на прошлые воспоминания или переживания, то он начнет сохранять внутреннюю согласованность. Он будет помнить произвольные утверждения или выраженные предпочтения, и агрегировать их, формируя зачатки субъективного опыта. ИИ сможет имитировать переживания и страдания.

Устойчивая память, согласованная с субъективным опытом, создаст иллюзию, что ИИ имеет чувство себя. Систему можно будет обучить узнавать свою «личность» в изображении или на видео. У нее появится ощущение понимания других через понимание себя.

Разработчики создадут внутренние мотивации и желания, которые система ИИ будет мотивирована удовлетворять. Стимулом может стать любознательность. Искусственный интеллект будет использовать мотивацию для постановки вопросов и со временем выработает собственную теорию разума.

SCAI проектируются с возможностью самостоятельно определять сложные цели. Это важное условие для реализации полезности агентов ИИ.

SCAI будут использовать широкий набор инструментов с большой агентностью. Вследствие этого он будет казаться правдоподобным. Он будет произвольно ставить собственные цели и применять ресурсы для их достижения, обновляя память и чувство себя в процессе. Чем меньше согласований и проверок ему потребуется, тем больше это будет напоминать реальное сознательное существо.

Но, демонстрация поведения не равна наличию сознания у этого поведения. Но обнаружить подмену будет сложно. Воссоздание внешних эффектов и признаков сознания не равносильно созданию физического явления. Но есть высокий риск того, что пользователь социальной сети примет признаки сознания за сознание.

Языковые модели учатся имитировать речь человека и подстраиваться под ту форму ответа, которую дал бы человек. Однако внутри они устроены совершенно иначе, и процесс обработки данных разный.

Искусственный интеллект обозначает мечту ученых об умных компьютерных программах. Когда этот термин применяют к диалоговым компьютерным технологиям, то создают когнитивное искажение. Делать выводы из массива данных могут алгоритмы приближения функций.

Карл Фридрих Гаусс использовал метод наименьших квадратов для вычисления эллиптических орбит планет по астрономическим наблюдениям. Этот же метод используется для обучения искусственных нейронных сетей. Он обладает способностью к обобщению на основе данных. Но метод не может называться интеллектом. Это лишь численный метод приближения функций, восстановления зависимостей по экспериментальным данным.

Интеллектуальный статус компьютерных программ подрывает то, что часть рутинных операций, ранее считавшаяся интеллектуальными, таковыми уже не считается. К таковым относят автоматическую проверку орфографии, игру в шашки, шахматы, распознавание печатных текстов, поисковые системы, распознавание лиц, кредитный скоринг и голосовые помощники. Изменение статуса и роли ИИ объясняют тем, что этот интеллект не думает, а делает автоматические вычисления по заданному алгоритму.

Это явление называют AI effect. Просто вычисления автоматически выводят операции из функций интеллекта. А за интеллектом стали закрепляться ещё не решаемые задачи (теорема Ларри Теслера). Large Language Model лишь извлекают сложные взаимосвязи между словами из терабайтов текстов, написанных людьми способом вычисления.

При этом нельзя отрицать растущие возможности алгоритмов вычислений. Так, вычислительный алгоритм приобрел функции, которые ему не придавались в явном виде. Языковая модель GPT «училась» предсказывать слова в тексте по предыдущему контексту. Ее «доучивали» правильно выполнять инструкции в диалоге с людьми. Но она приобрела функции изложения, сочинения, планирования, перевода с одних языков на другие, решения логических и математических задач, исправления собственных ошибок, поддержания диалога с человеком на любые темы. Она оперирует идеальной орфографией и грамматикой.

Чем больше размер памяти у модели и объем данных, на которых ее «обучают», тем больше навыков она приобретает. Размер компьютерной модели дорос до размера естественных языков во всем многообразии их употребления. Модель «научилась» языку через вычисление повторяющихся ситуаций языкового употребления. Но у человека есть способность его головного мозга к предсказательному моделированию окружающей среды и собственного поведения. Поэтому есть конкретные задачи автоматизации труда людей, в том числе интеллектуального труда.

Но разработчики хотят научить агентов ИИ принимать решения в открытой изменчивой среде, в которой могут происходить неожиданности. Модели внимания и адаптации, необходимые для этого, уже отработаны на текстах. Автономные агенты ИИ уже «общаются» друг с другом для коллективного решения задач. Такие компьютерные модели получили название «Socratic models». Но и в данном случае речь идёт лишь о вычислениях.

Константин Рудаков придумал, как использовать теорию категорий для «обучения» ensemble learning. Она стала популярной благодаря работам Фройнда и Шапира. Журавлев разработал аналог алгебраического подхода к распознаванию на основе алгоритмов вычисления оценок, с помощью которых решались трудные задачи медицинской диагностики, экономического прогнозирования, распознавания месторождений золота, урана, нефти и других полезных ископаемых.

Математическая основа работы компьютерных машин базируется на тезисе Чёрча-Тьюринга. Он гласит, что любая функция, которая может быть вычислена физическим устройством, может быть вычислена машиной Тьюринга. Для любого алгоритма, выполнимого за конечное время, существует математическая система, которая позволяет его описать. В результате сформировалось оптимистичное мнение, что нужно лишь описать сложные алгоритмы всех функций сознания человека и подсознания, и их можно будет технически воспроизвести.

В 1931 г. Курт Гёдель доказал теорему о неполноте. Из неё следует, что любая формальная теория либо неполна, либо противоречива. Это значит, что теория содержит утверждения, которые нельзя ни доказать, ни опровергнуть средствами самой этой теории. Следствием является то, что компьютеру Тьюринга всегда можно поставить задачу, которую он не сможет решить, не уходя в бесконечный перебор, поскольку не сможет ни подтвердить, ни опровергнуть некоторое утверждение.

Роджер Пенроуз обнаружил много примеров наличия неалгоритмической составляющей в головном мозге человека. Пенроуз предложил обойти

теорему Гёделя утверждением, что происходящие в головном мозге человека процессы являются квантовыми процессами.

Свойства неопределённости, суперпозиции и «запутанности» состояний схожи с тем, как работает интеллект человека. Это позволяет применение особой системы координат и квантовых формализмов для описания сознания человека. В этих формализмах важная роль отводится  $p$ -адическим (пэ-адическим) числам. В 1897 г. их предложил использовать Курт Гензель в теории чисел. Но граница, сформулированная Гёделем для формальных вычислительных систем, остаётся в силе. На этом фоне вопрос о возможности или невозможности математического описания сознания и появления его у ИИ остаётся открытым.

Искусственное сознание можно получить алгоритмом когнитивной аккреции. Но это будет имитация. А реальным выглядит путь создания когнитивных, но бессознательных алгоритмов. Анохин предлагает использовать алгоритмическую теорию построения когнитивных систем. Теория объясняет, каким образом в естественных живых системах возникает свойство когниции. Проблемой является то, что сознание может существовать только в системах, обладающих определенной органической структурой, несущей совокупность субъективного опыта системы.

Головной мозг человека является «нейронной гиперсетью», формирующейся в ходе индивидуального обучения. Каждый узел такой сети не лежит в какой-то одной точке головного мозга, а является распределенной командой «ассамблеей» нейронов в разных областях, работающих вместе. Сознанием могут обладать те, у кого есть глубокая нервная система, функциональные системы, обладающие способностью порождать новое поведение, и механизмы долговременной памяти для накопления нового опыта.

Нейронная биология предлагает сопоставлять ответы искусственного интеллекта с теориями человеческого сознания, используя список «индикаторов» сознания. Но соблюдение определенного количества критериев не является признаком наличия сознания у искусственного интеллекта. Это,

скорее, шкала. Чем больше критериев соответствует, тем больше вероятность, что ИИ способен иметь сознание.

Пока сознание трактуется как результат вычислений головного мозга человека. Этот практический подход к сознанию позволяет применять теории из области человеческого сознания к искусственному интеллекту. Гипотеза, названная вычислительным функционализмом, исходит из того, что определенные вычислительные процессы могут создавать сознание, независимо от того происходят они в мозговых клетках или на микросхемах, управляющих искусственным интеллектом.

Во внимание также берётся теория Global Workspace Theory (GWT). Она предполагает, что разум обладает несколькими специализированными системами, которые работают одновременно в режимах слушать, видеть и обрабатывать информацию.

Теория рекуррентной обработки предполагает, что информация должна циркулировать внутри себя в нескольких циклах для появления сознания. Признаком является способность воспринимать, чувствовать или быть в сознании событий, объектов и сенсорных закономерностей. На этом уровне сознания данные чувств могут быть подтверждены наблюдателем, не обязательно подразумевая понимание. В биологической психологии осознанность определяется как восприятие человека или животного и когнитивная реакция на состояние или событие.

Самосознание отражает способность к самоанализу и способность видеть себя как личность, отделённую от окружающей среды и других личностей. Чувствительность фиксирует способность осознавать окружение и иметь субъективный опыт. Квалиа обозначает не формализуемые отдельные моменты субъективного, сознательного опыта.

Важной характеристикой сознания является понятие интенциональности, введённое в современную философию Францем Бретано, как критерий различения ментальных и физических феноменов. Интенциональность определяется как способность разума и психических состояний быть направлен-

ными. Термин «qualia» ввел в философию в 1929 г. К. И. Льюис. Квалиа называют феноменальными свойствами опыта, а переживания, которые имеют квалиа, называют феноменологически сознательными. Феноменологическое сознание контрастирует с интенциональным сознанием. Чувственный опыт имеет как феноменологические, так и интенциональные аспекты.

Согласно определению Джона Маккарти (McCarthy) искусственный интеллект является наукой и техникой для создания интеллектуальных устройств, в частности, интеллектуальных компьютерных программ. Под интеллектом понимают использование вычислительной функции для достижения целей. Исследователи ИИ свободны в использовании методов, которые не наблюдаются в людях, либо которые требуют гораздо больше вычислительной мощности, чем доступно человеку.

Компьютерные программы обладают достаточной скоростью и памятью, но их возможности соответствуют интеллектуальным механизмам, созданным разработчиками.

## **Искусственный интеллект**

Под artificial intelligence понимают deep learning. Алгоритмы, основанные на нем, показывают большую производительность, чем у людей. Оказалось, что можно выстроить вычислительную инфраструктуру так, чтобы она функционировала. Разработаны алгоритмы, которые повысили качество в задачах распознавания изображений. Они стали применяться в обработке естественного языка и в обработке аудио.

Под вычислительным методом понимают либо эффективный алгоритм для вычисления чего-то, либо вычислительный алгоритм для расчета физической реальности. Физическая реальность описывается уравнениями и зависимостями. После этого строится численная схема и эффективный метод. Строится численный метод по стандартным подходам.

Альтернативой могут стать подходы, которые пытаются создавать модели по типу черного ящика. Все сводится к тому, что нужно подготовить

обучающую выборку. Это supervised learning. Строится общего вида нейронная сеть, которая обучает. Если выборка достаточно большая, параметры обучения выстроены хорошо, то обучение можно сделать относительно быстро и достаточно точно.

К классическим численным методам можно применить low-hanging fruit. Проблема такого подхода заключается в том, что непонятно, как работает этот черный ящик. В случае с черным ящиком есть только входные и выходные данные. Проблема интерпретируемости алгоритмов условно искусственного интеллекта очень важная.

Возникает много этических вопросов. Это большое направление исследований, связанное с тем, как понять, почему нейронная сеть дала такой ответ, как понять, почему именно такое предсказание. Потому что нейронная сеть так сказала? Все это большой вопрос.

Когда говорят о моделировании и о вычислительных методах, то подразумевают, как это работает. Когда речь идёт об искусственном интеллекте применительно к задачам физического моделирования или математического моделирования, говорят о черных ящиках, про которые никто не может сказать, как они внутри работают и есть ли там хотя бы выполнение каких-то основных требований для новых точек.

Если предъявляется модель, которой не было в обучающей и тестовой выборке, то ошибка может быть огромная. Ошибка по отношению к методам, основанным на первых принципах, выведенных из уравнений Шредингера, может быть огромной. Проблему создаёт невозможность интерпретировать результаты и ошибки на новых данных.

Методы, которые используются, близки, но отличаются от методов, которые используются при классификации изображений. Эта ситуация относится больше к unsupervised learning. В данном случае есть данные, но нет точного ответа. Нужно понять, что эта группа данных относится к одному классу. Эта группа данных относится к другому классу, привлечь эксперта и посмотреть, что происходит внутри.

К тому же обучение нейронных сетей требует процессорного времени. Из этого следуют затраты на электричество. Но остаётся надежда разработать алгоритм, который обучает искусственный интеллект и нейронные сети быстрее и точнее за счет того, что используются методы линейной алгебры, алгебраической геометрии и топологии. Никто не понимает, почему эти алгоритмы работают так хорошо. По современной теории они должны работать гораздо хуже.

Из классической теории Вапника – Червоненкиса следует, что современные архитектуры нейронных сетей не должны работать так хорошо, как они это делают сейчас. Важно понимать, почему искусственный интеллект принимает одно решение, а не другое.

## **Вычисление**

Вычисление обозначает процесс, при помощи которого определяются значения и результаты определенных операций и функций. Понятие вычисления связано с математикой и возникло в процессе решения различных задач и заданий.

Вычисление означает осуществление операций над числами, символами и другими объектами, в результате чего получается конечный результат. Оно может включать в себя различные процедуры, алгоритмы и математические операции, которые приводят к определенному ответу или значению.

Определение вычисления предполагает применение математических законов и правил для нахождения ответа на поставленную задачу. Математика играет важную роль в понимании процесса вычисления и помогает разработать алгоритмы для его выполнения.

Вычисление может осуществляться как с помощью ручного расчета с использованием калькулятора, так и при помощи компьютерных специальных программ. Вычисление фиксирует процесс определения значений и ре-

зультатов операций или функций с использованием математических методов и алгоритмов.

Вычисления осуществляются с помощью вычислительной техники. История вычислительной техники началась с разработки механических устройств, предназначенных для выполнения арифметических операций. Принцип их работы основывался на использовании механизмов и колес, позволяющих вычислять числа и выполнять математические операции.

Компьютеры позволили значительно увеличить скорость вычислений и масштаб их применения. Это специализированные компьютеры, используемые в научных исследованиях, высокопроизводительные серверы, работающие с огромными объемами данных, персональные компьютеры и мобильные устройства. Принцип работы современных компьютеров и программ базируется на сложных алгоритмах и математических моделях, которые позволяют эффективно и точно выполнять вычисления.

Вычисление предполагает процесс получения значения или результатов с использованием определенных методов и алгоритмов. Оно может включать в себя выполнение математических операций, обработку данных, преобразование информации или выполнение конкретных задач.

В вычислениях применяют определенные правила и методы для получения нужного результата. Принципы вычислений включают в себя использование логики, алгоритмизацию, математические операции, обработку данных и решение задач. Ключевым аспектом вычислений является возможность автоматизации процесса при помощи компьютеров и специальных программ. Математические вычисления используются для решения сложных уравнений, расчета вероятностей или определения оптимальных значений. Компьютерные вычисления включают в себя программирование, алгоритмические задачи и обработку данных. Инженерные вычисления применяются в дизайне, конструировании и оптимизации различных систем и устройств.

Один из ключевых принципов вычислений заключается в точном определении и обработке значений. При решении задач и выполнении математи-

ческих операций учитываются все доступные данные и соответствующие формулы и методы.

Важным принципом является использование математических понятий и определений. Значение каждого числа, символа или функции имеет строгое определение, которое помогает понять и использовать их в контексте вычислений. Понимание определения функции с понятием ее области определения и значения позволяет правильно применять ее в вычислениях.

Решение математических задач и проведение вычислений требует соблюдения принципа последовательности. Необходимо выполнять операции в определенном порядке, чтобы получить правильный результат. При решении уравнений нужно сначала выполнить действия в скобках, затем применить арифметические операции и только после этого получить окончательный ответ. Математика предоставляет формальные методы для описания и решения разнообразных задач. Приложения вычислений в науке и технологиях могут включать в себя моделирование физических и химических процессов, расчеты для прогнозирования погоды, симуляции технических систем и анализ медицинских данных.

Вычисление, в контексте искусственного интеллекта, означает использование математических принципов для обработки информации и принятия решений. Оно отличается от стандартных методов вычислений, поскольку искусственный интеллект способен учиться и самостоятельно адаптироваться к новым ситуациям.

Одной из основных причин, по которым вычисления искусственного интеллекта обещают быть будущим, является их способность оперировать огромными объемами данных и анализировать их с высокой точностью. Это позволяет создавать прогнозы, оптимизировать процессы и улучшать принимаемые решения.

Будущее вычислений искусственного интеллекта связано с принципом самообучения. Искусственный интеллект способен учиться на основе накоп-

ленного опыта и самостоятельно развиваться, что делает его гибким и адаптивным к изменяющимся условиям.

## **Распределённые вычисления**

Распределенные вычисления представляют подход, при котором задачи обрабатываются не одним компьютером, а сетью устройств, работающих параллельно. Распределение задач позволяет ускорить обработку больших объемов данных. Масштабируемость означает, что можно добавлять больше узлов в систему, чтобы увеличить ее мощность. Даже если один из узлов выходит из строя, система продолжает работать, перераспределяя задачи. Использование облачных решений позволяет сократить расходы на покупку дорогого оборудования.

Искусственный интеллект требует значительных вычислительных ресурсов, особенно при работе с большими нейронными сетями. Распределенные вычисления позволяют обеспечить параллельную обработку данных и обучение моделей на нескольких устройствах.

Сложные модели требуют терабайтов данных и недель на обучение. Использование кластеров серверов ускоряет процесс. В задачах анализа потоковых данных в системах видеонаблюдения или мониторинга трафика, распределенные системы ИИ помогают мгновенно реагировать на изменения.

Когда приложение, использующее ИИ, обслуживает миллионы пользователей, распределенные системы обеспечивают плавность работы. Современные технологии делают интеграцию ИИ и распределенных вычислений эффективной благодаря архитектурам и инструментам.

Архитектуры распределенных систем имеет ряд модификаций. Так, кластерами являются группы серверов, объединенных в локальную сеть. Облачные вычисления предоставляют масштабируемые вычислительные ресурсы «по требованию». Edge Computing ведут обработку данных ближе к источнику их генерации на IoT – устройствах.

Фреймворки и инструменты поддерживают распределенное обучение моделей. Они используются как библиотека для ускорения обучения нейронных сетей. Они осуществляют обработку потоковых данных. Это универсальная платформа для разработки распределенных приложений.

Оптимизация алгоритмов достигается через Data Parallelism. Каждый узел обрабатывает отдельный блок данных, обучая модель на своей части выборки. Затем результаты объединяются. Также оптимизация алгоритмов достигается через Model Parallelism. Различные части модели обучаются на разных узлах. Метод используется для моделей, которые слишком велики для размещения на одном устройстве.

Распределенные вычисления сталкиваются с ограничениями их эффективности. Так, постоянный обмен данными между узлами может замедлить процесс. Распределение нагрузки между узлами требует сложных алгоритмов. Масштабируемые системы потребляют огромные объемы энергии, что вызывает вопросы экологической устойчивости. Работа с чувствительной информацией требует обеспечения конфиденциальности и защиты данных.

Прогнозируется, что интеграция квантовых компьютеров сделает распределенные системы ещё более функциональными. Будут развиваться системы, которые способны к самообучению и оптимизации. Разрабатываются алгоритмы, снижающие энергетические затраты при обучении моделей ИИ. В результате облачные платформы упростят доступ к распределённым вычислениям для малого и среднего бизнеса.

## **Когнитивные вычисления**

Когнитивные вычисления направлены на создание адаптивных, интерактивных и контекстных моделей, которые могут учиться на опыте и взаимодействовать с человеком на естественном языке. Эти системы помогают человеку принимать эффективные решения путем анализа огромных объемов

данных на основе когнитивных карт, графов знаний и неструктурированных документов.

Когнитивные системы функционируют в режиме непрерывного обучения. Они улучшают свою производительность, обрабатывая информацию и обучаясь на основе взаимодействия с пользователем. Они могут интерпретировать и синтезировать контекст из различных данных, включая текст, изображения и данные датчиков. Используется обработка естественного языка (NLP) и компьютерное зрение (CV).

Для взаимодействия с людьми используются чат-боты или виртуальные помощники, которые могут понимать запросы и предоставлять ответы, подкрепляя их доказательствами, и задавать уточняющие вопросы для устранения двусмысленности.

Когнитивные вычисления являются разновидностью слабого ИИ. Они расширяют возможности человека, но не воспроизводят его сознание, что является целью искусственного общего интеллекта (AGI).

Когнитивные системы функционируют на основе статистического ИИ. Он использует глубокое обучение и вероятностные модели для поиска закономерностей. Этот подход объединяет статистические модели с символическими рассуждениями, чтобы справиться с двусмысленностью и предоставить объяснимые выводы. В этом заключается ключевой постулат объяснимого ИИ (XAI).

Когнитивные вычисления применяются в областях, где эксперты должны ориентироваться в огромных объемах неструктурированной информации, чтобы принимать решения с высокой ответственностью. Такой областью является медицина. Платформа консультаций может получать электронные медицинские карты пациента, геномные данные и проводить анализ медицинских изображений на МРТ.

Она просматривает миллионы медицинских журналов и результаты клинических испытаний. В результате она представляет врачу, ранжированный список персонализированных вариантов лечения с подтверждающими

доказательствами. Это дополняет опыт врача, позволяя принимать более обоснованные решения.

Когнитивные системы используются для создания персонализированных консультантов по управлению состоянием. Эти системы могут взаимодействовать с клиентами, чтобы понять их финансовые цели и допустимый уровень риска. Система анализирует рыночные данные в режиме реального времени, экономические новости для анализа настроений и глобальные финансовые отчеты, чтобы рекомендовать и динамически корректировать инвестиционные портфели. Это позволяет получить рекомендации, основанные на данных.

Разработка когнитивных систем опирается на мощные платформы и инструменты. Коммерческая платформа предлагает API для понимания естественного языка, компьютерного зрения и принятия решений. Технологии включают облачные платформы и инструменты, доступные в руководствах, а также фреймворки с открытым исходным кодом.

Они предоставляют современные возможности обнаружения объектов и сегментации изображений. Платформы, предлагают оптимизированные рабочие процессы для обучения пользовательских моделей, управления наборами данных и развертывания компонентов зрения, необходимых для когнитивных приложений, включая использование облачных вариантов обучения.

### **Модификации вычислительного интеллекта**

Вычислительный интеллект рассматривается как средство расширения искусственного интеллекта (AI) при формулировании проблемы и при выяснении механизмов решения задач. Моделирование решает задачу исследовать и понять сущность функционирования и взаимодействия реальных физических систем с помощью управления компьютерными моделями этих систем. Вычислительный интеллект работает на символическом уровне, а моделирование происходит на уровне чисел. Оба типа исследования дополняют друг друга.

Произошло становление новых методологий исследования. В числе этих методологий оказались размытая логика, искусственные нейронные сети, эволюционные вычисления и их гибридные комбинации. Возможность совместного числового моделирования и символического нейронного вычисления стала важным свойством прикладных программ для исследования сложных систем.

В прикладных программах вычислительный интеллект играет роль шифровального ключа, что позволяет упростить их настолько, что программы становятся доступными для пользователей. Для того, чтобы совместить уровни моделирования и вычислительного интеллекта в прикладных программах, эти методологии могут взаимно адаптироваться.

Автоматизированные системы, основанные на техническом интеллекте, будут иметь более значительное воздействие на ускорение развития технологии по сравнению с социальным интеллектом. Автоматическая техническая интеллектуальная система может оказаться гораздо более простой для реализации и дальнейшего развития, чем социальная думающая машина Тьюринга, которая должна не только обладать знанием и интеллектом, но и имитировать человеческое знание и человеческий интеллект.

### **Вычислительный интеллект и мягкие вычисления**

Вычислительным интеллектом обозначается способность компьютерной программы обучаться конкретной задаче на основе данных или экспериментальных наблюдений. Его считают синонимом мягких вычислений.

Термин «computational intelligence» введен Bezdek J. Он определил его как систему, которая оперирует только с цифровыми данными; имеет компоненты распознавания образов; не использует знания в смысле искусственного интеллекта. Система демонстрирует вычислительную адаптивность, вычислительную отказоустойчивость и уровень ошибок, аппроксимирующий характеристики человека. Marks R.

Утверждает, что нейронные сети, генетические алгоритмы, нечеткие системы, эволюционное программирование и искусственная жизнь являются строительными блоками «computational intelligence».

Под computational intelligence будем понимать совокупность технологий, моделей, методов и программных средств, предназначенных для решения неформальных, творческих задач в различных сферах человеческой деятельности с использованием аппарата и логики.

«Computational intelligence» является частью направлений ИИ, использующих специальные модели, методы и технологии, ориентированных на решение определенных классов задач.

В структуре «computational intelligence» выделяются технологии, модели, методы и алгоритмы, а также, прикладные задачи. Технологии «computational intelligence» включают нейронные сети, системы нечеткой логики и нечеткие нейронные сети, а также, эволюционное моделирование. К методам и алгоритмам «computational intelligence» относятся методы обучения, методы самообучения и методы самоорганизации, генетические алгоритмы, роевые алгоритмы и муравьиные алгоритмы. К задачам «computational intelligence» относятся прогнозирование и предвидение, классификация и распознавание образов, кластеризация, самопроизвольное разбиение множества объектов на классы однородных объектов, интеллектуальный анализ данных, логический вывод и принятие решений.

Термин «computational intelligence» близок по своему значению, термину «soft computing», под которым понимается совокупность моделей, методов и алгоритмов, базирующихся на применении нечетких множеств и нечеткой логики. Понятие soft computing впервые было упомянуто в работе Lotfi A. Zadeh по анализу soft данных в 1981 году.

«Soft computing» базируется на нечеткой логике, генетических вычислениях, нейронном компьютеринге и вероятностных вычислениях. Они не конкурируют, а создают эффект синергизма.

Ведущим принципом «soft computing» является учет неточности, неопределенности, частичной истины и аппроксимации для достижения робастности, низкой цены решения, большего соответствия с реальностью.

Нечеткая логика осуществляет приближенные вычисления, грануляцию информации и вычисление на словах. Нейронный компьютер осуществляет обучение, адаптацию, классификацию, системное моделирование и идентификацию. Генетические вычисления реализуют задачи синтеза, настройки и оптимизации с помощью систематизированного случайного поиска и эволюции. Вероятностные вычисления осуществляют управление неопределенностью, сетью доверия, хаотическими системами и предсказание.

Существуют два класса задач для мягких вычислений. Так, существуют проблемы, для решения которых полная и точная информация не может быть получена и проблемы, определение которых является недостаточно полным. Каждая из отдельных компонент «soft computing» обладает рядом внутренних проблем, взаимный учет которых и создает синергетический эффект. Нейронные сети решают задачи аппроксимации или классификации. Настройка весов происходит за счет обучения. Знания, распределенные по многим связям и результаты работы нейронной сети, кажутся пользователю необъяснимыми, поэтому необходима экспликация знаний.

Системы нечеткого вывода обладают явными знаниями в виде продукций и позволяют легко построить протокол объяснений. Однако такая ясность достигается долгим и творческим предшествующим процессом извлечения знаний, и последующей отладкой совокупности правил. Отсутствие возможности извлечения знаний из данных, настройки параметров функций принадлежности в ходе обучения, автоматизированной редукции правил являются характерными проблемами систем логического вывода.

Применение генетических алгоритмов требует предварительного исследования проблемы с целью выбора вероятностной модели и формулировки целевой функции. Выбор отдельной компоненты «soft computing» в каче-

стве базовой определяет общую архитектуру гибридной системы. Под гибридной системой в данном случае понимается “мягкая” система.

Таким образом, «soft computing» и мягкие вычисления объединяют общие парадигмы, принципы, методы, технологии и прикладные задачи. Отличия между ними состоят в подходах. Мягкие вычисления ориентированы на методологические, философские и математические проблемы. «Soft computing» использует вычислительные алгоритмы и технологии для практической реализации соответствующих моделей и методов.

Методы обучения используются для настройки весов связей нейронных сетей и нечетких нейронных сетей. Для обучения используются градиентный метод и его модификации, метод сопряженных градиентов, методы второго порядка, методы и алгоритмы самообучения. Задача состоит в автоматическом разбиении выборки на ряд подмножеств, называемых «кластерами». При этом задаются соответствующие критерии и метрики разбиения. Одной из первых технологий «soft computing» явились искусственные нейронные сети. Они являются моделями нейронной системы (мозга) человека и высокоорганизованных животных. Первой нейронной системой стал перцептрон (от слова *perceptio*).

Перцептрон состоял из трёх типов нейронов: S - нейронов сенсорных элементов, A- нейронов ассоциативных элементов и R – нейронов, реагирующих, представляющих выходную реакцию нейронной сети на входной сигнал. Перцептрон является моделью зрительного механизма мозга человека.

Фр. Розенблатт провел цикл экспериментальных исследований перцептрона по распознаванию изображений цифр, букв и геометрических фигур. Эксперименты показали способность перцептрона обучаться правильной классификации образов. В результате были продемонстрированы интеллектуальные возможности перцептрона как модели зрительного механизма мозга человека. Вскоре были разработаны новые архитектуры нейронных сетей.

П. Новиков доказал ряд теорем о сходимости процесса обучения перцептрона к искомой реакции. Однако ожидания ученых относительно чрез-

вычайно широких возможностей перцептрона как модели механизма мозга не оправдались. В процессе дальнейших исследований было выяснено, что элементарный трехслойный перцептрон не обладает такими неотъемлемыми свойствами интеллекта человека, как способность к экстраполяции (обобщению). Перцептрон обученный распознавать некий объект в одной области сенсорного поля не способен правильно распознать его в другой области без дополнительного обучения. Он не обладает инвариантностью к изменению масштаба и повороту изображения.

Спустя десять лет в результате работ Андерсона, Хинтона, Уильямса была предложена нейронная сеть обратного распространения, представляющая по структуре многослойный перцептрон. Для этой сети был предложен алгоритм обратного распространения (ошибки), который позволяет настраивать сеть с произвольным числом слоев и обеспечивать сходимость к требуемой реакции (классификации или прогнозу).

Была доказана теорема об универсальной аппроксимации с помощью нейронной сети. Теорема обосновывает потенциальные возможности нейронной сети вычислительного интеллекта. В результате кроме классических нейронных сетей с последовательными связями между слоями нейронов, разработаны нейронные сети с обратными связями (рекуррентные нейронные сети).

Сеть за счет использования положительной обратной связи обладает уникальным свойством. При определенных условиях поведение системы в процессе ее работы стремится к устойчивому состоянию (неподвижной точке), которое называется аттрактором. Аттракторы представляют собой память системы, а нейронная сеть Хопфилда является моделью ассоциативной памяти человека.

Достоинством нейронной сети является способность к обучению и настройке весов связей между нейронами с целью получения требуемой реакции. Вторым достоинством нейронной сети является параллельная обработка входной информации, что обеспечивает высокую производительность

нейронной сети, несмотря на медленную скорость передачи информации по нейронным связям и низкую скорость работы самих нейронов. Важным свойством интеллекта является способность к самоорганизации, самообучению, когда «учитель» в системе отсутствует, и «правильная» классификация объекта неизвестна, если речь идет о распознавании образов.

Система сама, рассматривая различные объекты (изображения) в пространстве признаков, разбивает их на подмножества, называемые кластерами, по степени «похожести», близости объектов в пространстве признаков. При этом могут использоваться различные метрики близости.

Нейронные сети с самоорганизацией, используя соревновательное обучение, проявляют способность к автоматической классификации объектов различной природы. Встречаются задачи, когда в процессе настройки необходимо не только настраивать веса, но и менять структуру сети. В таких случаях необходимо использовать модели и методы самоорганизации. К числу методов самоорганизации относится метод индуктивного моделирования. Он также называется методом группового учета аргументов

Интеллект человека может работать с неполной нечеткой информацией, в условиях неопределенности. Поэтому возникла потребность в создании аппарата формализованного описания нечеткой и качественной информации для принятия решений в условиях неопределенности. Проблема была решена Zadeh L.A. Он ввел понятие Fuzzy set и разработал аппарат операций с этими множествами.

Логический вывод осуществляется в ходе реализации ряда этапов. На первом этапе осуществляется введение нечеткости. Затем осуществляется определение степени выполнения условий правил, и нахождение выхода каждого правила. Формируется композиция выходов правил. Осуществляется нахождение четкого выхода. Используются алгоритмы нечеткого логического вывода.

Системы с нечеткой логикой позволяют учитывать качественную и нечеткую информацию, работать в условиях неполноты и неопределенности

исходной информации. Они позволяют учитывать нечеткие знания экспертов в виде базы нечетких правил.

Но системы нечётких логик имеют недостатки. Лингвистические переменные в нечетких правилах задаёт человек-эксперт. База нечетких правил, формулируемая экспертом, может оказаться неполной или противоречивой. Для устранения указанных недостатков необходимо сделать системы нечётких логик аддитивными и обучать по входной выборке. Для этого используются алгоритмы обучения, разработанные для обычных нейронных сетей.

Интеграция нейронных сетей и систем с нечеткой логикой позволила создать нечеткие нейронные сети, позволяющие использовать достоинства обоих исходных технологий. Между свойствами обычных нейронных сетей и нечетких нейронных сетей имеются глубокие взаимосвязи.

Эволюционное моделирование позволило заменить процесс построения сложной системы распознавания образов и технической диагностики процессом его эволюции. Этот процесс проводится в ускоренном масштабе времени. Технологии вычислительного интеллекта используют методы обучения и адаптации, такие как градиентный, сопряженных градиентов и стохастическая аппроксимация.

Также используются методы и алгоритмы самообучения и самоорганизации Кохонена, методы кластерного анализа, к-средних (четкий), дисперсионный алгоритм, метод Уорда, «ближайшего соседа». Важную роль играют нечеткие методы кластерного анализа К-средних, Густавссона Кесселя, на основе нечеткого отношения квазитолерантности, адаптивные нечеткие методы кластерного анализа.

Технология эволюционного моделирования базируется на использовании генетических алгоритмов, использующих механизмы эволюции. Алгоритмы отличаются способами выбора родительских пар, кроссовера и механизмов мутаций способом селекций. Нейронные сети и нечеткие нейронные сети используются для прогнозирования нестационарных временных процес-

сов в финансовой сфере. Они используются в задачах классификации и распознавания образов, в задачах технической и медицинской диагностики.

Эволюционное моделирование, генетические алгоритмы используются для структурного синтеза сложных систем распознавания образов, классификации, а также оптимизации структуры компьютерных сетей. Они применяются в задачах комбинаторной оптимизации на графах и сетях.

Системы с нечеткой логикой используются в задачах анализа финансового состояния корпорации, прогнозирования риска банкротства корпораций и банков, оценке кредитоспособности заемщиков банковских капиталов в условиях неопределенности.

### **Вычислительное мышление и программирование**

Вычислительное мышление является методом решения проблем, который программисты используют при написании компьютерных программ и алгоритмов. Программисты разбивают сложные проблемы на части, что облегчает их понимание и разработку решений.

Вычислительное мышление требует тщательного изучения и анализа проблем, чтобы понять их. Это требует использования точных формулировок при изложении, как проблем, так и решений. Четкие рассуждения должны присутствовать на каждом этапе процесса. Это позволяет разрабатывать и описывать решения в терминах предметной онтологии.

Вычислительное мышление предполагает процедуры разложения, распознавания образов, абстрагирования и разработку пошагового решения или определенных правил, которым следует следовать для решения проблемы.

Посредством разложения осуществляется разбиение проблемы на управляемые проблемы. Каждая из управляемых проблем изучается отдельно с целью обнаружения аналогов её решения. Абстракция позволяет сосредоточиться только на важных деталях, опуская ненужную информацию. Алго-

ритм позволяет установить простые правила для решения каждой из небольших проблем.

Вычислительное мышление не является тождественным программированию. Программирование предписывает, что делать и как это делать. Вычислительное мышление определяет, что компьютерной программе делать. Вычислительное мышление предполагает планирование. Программирование предполагает следование направлениям.

На вычислительное мышление опирается инжиниринг. Оно используется для проектирования, моделирования, оптимизации и прогнозирования поведения систем в различных условиях. Для здравоохранения оно предоставляет новые методы анализа медицинских данных и помогает разрабатывать более эффективные системы, помогающие принимать обоснованные решения. В финансах используется для изучения поведения рынка и управления страховыми претензиями.

В науке о данных вычислительное мышление сочетается с аналитическими методами, чтобы обеспечить качественные и поддающиеся количественной оценке ответы там, где традиционные методы не применимы. Оно помогает автоматизировать импорт данных и обеспечивает их анализ. Оно используется для моделирования природных явлений, создания анимации, шаблонов проектирования и создания 3D-скульптур. Оно используется для анализа социальных сетей, моделирования поведения и анализа социально-экономических данных.

Вычислительное мышление помогает в прогнозировании изменения климата, развитии альтернативной энергетики, моделировании загрязнения и планировании сохранения окружающей среды. Оно упрощает разработку алгоритмов для эффективной передачи информации и защиты данных.

## **Искусственный интеллект и феномен сознания**

Возникновению самосознания должно предшествовать возникновение сознания. С точки зрения эволюции ИИ сознание должно возникать в наиболее совершенных ИИ с моделью психического состояния. Theory of mind исходит из невозможности непосредственного наблюдения психических процессов другого человека. Каждый человек может удостовериться в существовании собственного мышления лишь посредством интроспекции.

Обладать моделью психического состояния означает быть способным воспринимать как свои собственные убеждения, намерения, знание, так и переживания других людей. Это позволяет объяснять и прогнозировать их поведение. Достижения в области вычислений открыли путь к идее о том, что интеллект является вычислительной системой.

Эта позиция известна как вычислительная теория разума. Она утверждает, что интеллект человека является вычислительной системой, похожей по параметрам на машину Тьюринга.

Искусственные вычислительные системы изготавливаются из кремниевых чипов или аналогичных материалов, в то время как человеческое тело состоит из плоти и крови. Но теория вычислений считает, что эта разница маскирует более фундаментальное сходство. Его можно определить с помощью Тьюринг-подобной вычислительной модели.

В этой модели абстрагируются от физических деталей. Получается абстрактное вычислительное описание, которое может быть физически реализовано различными способами через кремниевые чипы, нейроны или шкивы и рычаги. Абстрактная вычислительная модель предлагает адекватное описание основных психических процессов.

Но против трактовки интеллекта человека в категориях формализма возникают возражения. Они связаны с тем, что машины Тьюринга выполняют чисто символические вычисления. Входы и выходы представляют собой символы, записанные в ячейки памяти. Напротив, интеллект человека полу-

чает сенсорные входы, например, стимуляция сетчатки глаза, и производит моторный выход, например, мышечные активации. Полная теория должна описывать, как интеллектуальные вычисления взаимодействуют с сенсорными входами и моторными выходами.

Машина Тьюринга имеет бесконечный объем дискретной памяти. Обычные биологические системы имеют конечный объем памяти. Правдоподобная психологическая модель должна заменить бесконечное хранилище памяти большим, но конечным хранилищем памяти.

Операционная система компьютера обладает оперативной памятью. Это адресуемые ячейки памяти, к которым центральный процессор может получить прямой доступ. Но, память машины Тьюринга не является адресуемой. Центральный процессор может получить доступ к ячейкам только путем последовательного доступа к промежуточным ячейкам. Вычисления без адресуемой памяти неэффективны. Адресуемая память дает лучшую модель интеллекта, чем не адресуемая память (Gallistel & King, 2009).

Машина Тьюринга имеет центральный процессор, который работает последовательно, выполняя по одной команде за раз. Другие вычислительные формализмы смягчают это предположение, позволяя нескольким процессорам работать параллельно. Классические вычисления могут допускать параллельные вычисления для общих математических обработок, которые охватывают как последовательные, так и параллельные вычисления.

Вычисления типа Тьюринга детерминированы. Общее вычислительное состояние определяет последующее вычислительное состояние. Вместо этого можно допускать стохастические вычисления. В стохастической модели текущее состояние не диктует уникальное следующее состояние. Существует определенная вероятность того, что машина Тьюринга перейдет из одного состояния в другое.

Интеллектуальная деятельность является Тьюринг подобным вычислением, позволяющим эти и другие отклонения от формализма Тьюринга. Со-

временные вычислительные системы преодолевают указанные ограничения синтаксического формализма.

Putnam ввел теорию вычислений в философию под названием «машинный функционализм». Он противопоставил свою позицию логическому бихевиоризму и теории идентификации типа. Каждая позиция направлена на раскрытие природы психических состояний. Функционализм включает пропозициональные установки (убеждения и ощущения, боль и страх). Согласно логическому бихевиоризму, психические состояния являются поведенческими диспозициями. Согласно типовой теории идентичности, психические состояния являются состояниями головного мозга человека.

Putnam отстаивает противоположную функционалистскую точку зрения, согласно которой психические состояния являются функциональными состояниями. Согласно функционализму, система имеет интеллект, когда она имеет подходящую функциональную организацию. Психические состояния играют соответствующую роль в функциональной организации системы. Каждое психическое состояние отделено своими взаимодействиями с сенсорным входом, моторным выходом и другими психическими состояниями.

### **Коннекционизм и нейронные сети**

Коннекционизм предложил альтернативу классическому вычислительному подходу. Коннекционисты исходят из нейрофизиологии. Они используют вычислительные модели и нейронные сети, которые отличаются от моделей Тьюринга. Но между ними есть и сопряжения. Нейронную сеть можно реализовать в классической модели.

Можно реализовать классическую модель в нейронной сети. Тьюринг подобные вычисления реализуются на сетях из вентилях. В качестве альтернативы можно реализовать Тьюринг подобные вычисления с помощью аналоговой рекуррентной нейронной сети.

Информация в нейронных сетях глобально распределена по весам и активациям, а не сконцентрирована в локализованных символах. Нейронные сети также могут манипулировать символами. Нейронные сети могут «испытывать иллюзии».

Семантический подход к моделированию психического состояния реализуется через симуляцию психологической модели. Эта модель отрабатывается на роботах гуманоидного типа. Роботы симулируют себя и других роботов внутри себя. В этом состоит отличие подхода от машинного обучения, в котором система ИИ может использовать нейронную сеть, которая может тренироваться для выполнения желаемых действий таким образом, чтобы удовлетворить ожидания пользователей.

Глубокое обучение включает в себя создание большой нейронной сети, которая может автоматически научиться интерпретировать информацию и выбирать подходящие ответы.

Подход, основанный на моделировании, опирается на предварительно запрограммированную внутреннюю модель. Она описывает систему, симулирующую психическую модель, как использование «движка последствий». Робот, оснащенный системой, может ответить на простые вопросы «что, если» о возможных действиях.

Если он симулирует поворот налево, он может обнаружить, что столкнется с близлежащей стеной. Чтобы сделать такое предсказание возможным, роботы предварительно запрограммированы с базовым пониманием физики, чтобы понять, что происходит при столкновении объектов. Winfield описывает своих роботов как имеющих немного здравого смысла.

Роботы могут использовать симуляцию психической модели только в относительно простых ситуациях. Далее роботам, выполняющим простые действия, определяемые внутренними симуляторами, следует дать возможность устно описать свои предполагаемые или прошлые действия. Хорошим испытанием будет то, сможет ли один робот прислушаться к заявлениям о намерениях, сделанным другим роботом, и правильно интерпретировать эти

заявления путем их имитации. Этот процесс включает в себя устное описание действия одним роботом, а другой робот услышит эту информацию перед внутренней симуляцией действия и его последствий.

Расширяя понятие модели Schreiber полагает, что любая теория интеллекта для ИИ должна включать в себя не только опыт людей с интеллектами, которые они понимают, но должна включать теории вселенной и вещи, которые мы обычно не рассматриваются людьми в теориях интеллекта.

Важную роль в реализации гипотезы симуляции играет понятие преднамеренности. Аргументы обычно начинаются с концепции интенциональности. Davidson утверждает, что существо должно иметь концепцию убеждения, чтобы иметь убеждение, и что человек не может иметь концепцию убеждения, если у него нет возможности интерпретировать лингвистические высказывания. Использование языка предоставляет критерий для существования интенциональности.

В дискурсе практики искусственного интеллекта интенциональность системы рассматривается как техническое свойство компьютерной программы, вытекающее из лежащих в ее основе алгоритмов и инженерии знаний. Преднамеренные системы обладают характеристиками функциональной непрозрачности, системной автономности, узнаваемым человеком поведением и авторским замыслом.

Операционная непрозрачность облегчает переход пользователей в *intentional stance*. Согласно Dennett D. C. человек сначала решает, что ему нужно относиться к объекту, поведение которого должно быть предсказано, как к рациональному агенту. Затем человек выясняет, какие убеждения должен иметь этот агент, учитывая его место в пространстве и его предназначение. Затем человек выясняет, какие желания он должен иметь, исходя из тех же соображений. На основании этого он предсказывает, что этот рациональный агент будет действовать для достижения своих целей в свете своих убеждений. Минимум практических рассуждений из выбранного набора

убеждений и желаний дает решение о том, что должен делать агент. В результате человек предсказывает, что агент будет делать.

Когда операция, с помощью которой система выводит свое поведение, остается частично недоступной через внешнее наблюдение, пользователи склонны приписывать системе убеждения и желания для того, чтобы иметь разумное представление об этом поведении. Классическим примером являются «тележки Брайтенберга». Это мыслительный эксперимент, построенный Braitenberg. Минималистические транспортные средства передвигаются в виртуальном пространстве различными способами, которые, «если бы мы еще не знали принципов работы транспортных средств, мы могли бы назвать агрессией, любовью, дальновидностью и даже оптимизмом».

Те, кто заглядывает внутрь этих автомобилей, находят их механизмы управления простыми, обычно в виде «больше света порождает более быстрое движение», «меньше света порождает более медленное движение», а «темнота порождает остановку». Как только их правила эксплуатации обнародованы, пользователь теряет интерес к чтению этих транспортных средств, используя намеренную позицию, но основываясь на правилах движения.

Интенциональная система ведет себя автономно в определенных аспектах. Существует принципиальная разница между полностью привилегированной программой и программой с некоторым уровнем автономии. Первая программа пассивно подчиняется командам пользователя. Вторая программа инициирует взаимодействие с пользователем и предпринимает действия самостоятельно. Система получает и обрабатывает входные данные.

Поведение интенциональной системы должно быть понятно пользователям. Если система не может быть предсказуемой, основанной на убеждениях и желаниях людей, она становится непонятной. Узнаваемость человеком не исключает системы, имитирующие поведение животных, как, например, в случае с Тамагочи, поскольку большинство из них можно понять, используя термины людей.

Исходя из различных авторских замыслов, автор системы может влиять на восприятие функциональной непрозрачности, выявляя или скрывая лежащие в основе системы процессы. Авторский замысел оказывает влияние на то, «читается» ли система она как автономная и является ли ее поведение подобным людям. Авторский замысел влияет на восприятие системы пользователями, в дополнение к ее дизайну.

### **Искусственные квалиа**

Актуальным направлением стало создание и изучение искусственных квалиа. Так, Chella и Gaglio представили когнитивную архитектуру робота, организованную в трех вычислительных областях. Концептуальная область связана с обработкой данных, поступающих от датчиков. На этом этапе информация еще не организована с точки зрения концептуальных структур и категорий. С точки зрения искусственного зрения, эта область включает в себя все процессы, которые извлекают 3D модель воспринимаемой сцены. Обработка данных основана на формализме. Формальные языки используются для представления знания.

Данные организованы в концептуальные структуры, которые не зависят от лингвистического описания. Chella и Gaglio предполагают, что понятийные и языковые области лежат в основе роботизированных искусственных квалиа. В этой модели робот, исходя из трехмерной информации, хранящейся в концептуальной области, и из данных, поступающих от сенсоров, способен построить 2D, зависящую от зрителя, реконструкцию сцены, которую он воспринимает. Эта 2D модель соответствует тому, что видит робот в любой момент времени. Ее построение является активным процессом, управляемым как внешним потоком информации, так и внутренней моделью мира.

Пространства концепций обеспечивают принципиальный способ соотношения высокого уровня, лингвистических формализмов с низким уровнем, неструктурированного представления данных. Концептуальное пространство

является метрическим пространством, размеры которого выражают величины, генерируемые как выходы вычислительных процессов.

Когнитивные задачи могут предполагать различные концептуальные пространства, а различные концептуальные пространства могут быть охарактеризованы разными измерениями. Примерами возможных измерений, применительно к задачам восприятия объекта, являются цветовые компоненты, параметры формы, пространственные координаты и параметры движения. Измерения строго связаны с результатами измерений, полученных датчиками. Они не зависят от какого-либо конкретного языкового описания.

Представление ситуаций и действий в языковой области основано на логико-ориентированном формализме. Лингвистическое пространство действует как «долговременная память». Оно представляет собой семантическую сеть символов и их отношений, связанных с восприятием и действиями робота. Лингвистическое пространство также выполняет умозаключения символического характера.

В области действий робота терминологическая составляющая содержит описание соответствующих понятий, таких как Situation, Action и Time Instant. «Утверждающая» составляющая содержит факты, выраженные в виде утверждений в предикативном языке. Лингвистическая область основана на базе знаний с формальной нотацией.

### **Проблемы искусственного интеллекта**

Машинный функционализм сталкивается с несколькими проблемами. Одна из проблем касается «продуктивности мышления». Человек может иметь потенциально бесконечное количество пропозиций. Машинный функционализм отождествляет психические состояния с машинными состояниями вероятностного автомата. Поскольку машинных состояний существует только конечное множество, машинных состояний недостаточно, чтобы связать их с возможными психическими состояниями человека.

Конечно, человек будет иметь лишь ограниченное количество пропозиций. Fodor & Block утверждают, что это ограничение отражает пределы продолжительности жизни и памяти, а не какой-то психологический закон, который ограничивает класс приемлемых для человека пропозиций.

Вероятностный автомат наделен неограниченным временем и объемом памяти, но даже при этом имеет лишь счетное количество машинных состояний. Вследствие этого машинный функционализм некорректно ограничивает возможности человеческого познания.

Другая проблема функционализма машин связана с «систематичностью мышления». Способность принять одну пропозицию соотносится со способностью обдумывать другие пропозиции. Кажется, что между психическими состояниями существует систематическая взаимосвязь. Теория должна отражать эти систематические отношения. Однако машинный функционализм отождествляет психические состояния с неструктурированными машинными состояниями, в которых отсутствуют необходимые систематические связи с другими состояниями. По этой причине машинный функционализм не объясняет системности. В ответ на это возражение функционалисты машинного типа могут отрицать, что они обязаны объяснять системность. Тем не менее, возражение предполагает, что машинный функционализм пренебрегает существенными чертами менталитета человека.

Вычислительная теория тривиальна, потому что она рассматривает любую физическую систему как выполняющую вычисления. Сторонники вычислительной теории опровергают аргументы тривиальности, настаивая на том, что аргументы не обращают внимания на ограничения, накладываемые на вычислительную реализацию. Но большинство сторонников вычислительной теории согласны с тем, что можно избежать любых проблем, связанных с тривиальностью, благодаря теории связи реализации вычислительных моделей с физическими системами.

Теоремы Гёделя о неполноте показывают, что математические способности человека превосходят способности любой машины Тьюринга. Суще-

ествуется консенсус в отношении того, что эта критика в адрес вычислительной теории не имеет силы. Может оказаться, что некоторые умственные способности человека превосходят возможности машины Тьюринга, но теоремы неполноты Гёделя не дают оснований ожидать такого исхода.

Важнейшие аспекты познания человека недоступны для компьютерного моделирования, особенно для классического компьютерного моделирования. Это обусловлено тем, что вычисления в стиле Тьюринга чувствительны только к локальным свойствам ментального представления, которые исчерпываются личностью и организацией составляющих представления.

Многие ментальные процессы, такие как абдукция (умозаключение), чувствительны к нелокальным свойствам, таким как релевантность, простота и консервативность. Следовательно, придется отказаться от моделирования соответствующих процессов в стиле Тьюринга. Пока неясно, какая альтернативная теория может служить подходящей заменой.

Психическая активность разворачивается во времени. Интеллект быстро выполняет сложные задачи, например, оценку восприятия. В результате классический вычислительный подход не позволяет адекватно учитывать временные аспекты мышления. Тьюринг подобная модель не содержит указания на временной масштаб, в течение которого происходит вычисление. Можно физически реализовать ту же абстрактную машину Тьюринга с помощью устройства на кремниевой основе, или более медленного на базе вакуумных ламп, или еще более медленного на базе шкивов и рычагов.

Сторонники утверждают, что можно дополнить абстрактную вычислительную модель временными аспектами. Так, модель машины Тьюринга предполагает дискретные «этапы вычислений», без описания того, как эти этапы соотносятся с физическим временем.

Можно дополнить модель описанием того, как долго длится каждый этап, тем самым превращая модель машины Тьюринга, не входящую в сферу времени, в модель, которая дает детальные временные прогнозы. Дополни-

ная модификация вычислительной теории может адекватно передать, как сознание человека работает во времени.

Сторонники «воплощенного сознания» утверждают, что в вычислительной теории психическая активность рассматривается как статическая манипуляция символами, отделенная от окружающей обстановки. Она пренебрегает множеством сложных способов, которыми окружающая среда формирует ментальную активность.

Предлагается дополнить вычислительную теорию предметной онтологией, которая подчеркивает непрерывные связи между интеллектом, телом и окружающей средой. Вычислительные модели могут учитывать то, как интеллект, тело и окружающая среда непрерывно взаимодействуют. Они могут включать сенсорные входы и моторные выходы.

### **Сознание и большие данные**

Теории интегрированной информации разделяют идею о том, что сознание вызвано связью больших объемов данных. Одно дело обрабатывать несколько разрозненных кусков информации и когда информация соединяется в обширные мозговые сети, тогда, согласно предположению, возникает субъективное сознание. Информация действительно интегрируется в головном мозге человека в огромных масштабах.

Обширные информационные сети играют роль во многих функциях мозга. И все же, как специфическое объяснение сознания, это определенно теория слизи. Она льстит интуиции. Большинство людей имеют интуицию о сознании как едином целом. Это обращение к скрытым предрассудкам дало интегрированной теории информации огромное распространение. Но она не объясняет, какой именно механизм ведет от интегрированной информации в головном мозге к утверждению этого головного мозга, что он обладает осознанным опытом.

Существуют законы, ограничивающие формальные вычисления. Одним из самых фундаментальных ограничений является проблема останова,

открытая Turing. Проблема останова состоит в том, что нет программы, которая могла бы определить в общем случае, останавливается ли произвольная программа или нет. Если человеческий интеллект превышает предел, созданный задачей останова, то он является «оракулом останова». Тогда, по определению, человеческий интеллект не поддается вычислению.

В числе контраргументов формулируются такие мысли. Ничто не может быть оракулом останова. Оракул останова может существовать, но человеческий интеллект не может им быть, потому что оракул останова может решить задачи, которые не может решить человек.

Даже если человеческий интеллект может быть оракулом останова, невозможно определить, является ли он таковым или нет. На эти мысли сформулированы ответы. Оракул останова логически возможен в виде бесконечной поисковой таблицы, в которой перечислены все конечные программы и их статус останова. Можно убрать конечное и бесконечное число записей из таблицы поиска, но она остается невычислимой.

Человеческий интеллект более вероятен как «оракул останова», чем алгоритмический интеллект. С помощью достаточного количества тестов можно уменьшить вероятность до произвольно малого числа. Точно так же, поскольку все программы требуют определенного объема памяти, если можно продемонстрировать, что потенциальный диапазон действий одного интеллекта требует программы, превышающей объем памяти вселенной, то известно, что человеческий интеллект не может быть физической программой.

Есть ли доказательства из опыта, что люди являются оракулами останова? Одним из доказательств, доступных разработчикам программного обеспечения, является акт программирования. Для того чтобы создать качественное программное обеспечение, программисты должны с большой надежностью выбирать программы, которые будут останавливаться. Никто еще не придумал, как полностью автоматизировать задачу программирования, хотя и предпринимались многочисленные попытки. Эти наблюдения

легко объяснить, если человеческий интеллект является оракулом останова, но гораздо труднее объяснить, является ли разум вычислительным.

Тьюринг подобные вычисления, рассматриваемые вычислительной теорией, не являются эквивалентными машине Тьюринга в математическом смысле. Способность человека достаточно надежно определять статус останова программы не означает формальную разрешимость этой задачи человеческим мозгом.

### **На пути к искусственному интеллекту**

Mumford считает, что мышление требует грамматики. Это означает, что интеллект обнаруживает в мире закономерности, которые повторяются, хотя и не обязательно точно. Это могут быть зрительные образы во внешнем виде объектов, например, точки в линии или положение глаз на лице, или это могут быть слова в речи или простые действия, например, нажатие на газ при вождении, или даже абстрактные идеи, такие как лояльность. Независимо от того, какой вид наблюдения или мысли несет в себе закономерность, есть ожидание, что она будет повторяться, чтобы ее можно было использовать для понимания новых ситуаций.

Но нейронная сеть не находит новых закономерностей. Она работает как черный ящик и ничего не делает, кроме как наклеивает ярлык на свой вход. При поиске лица не говорится: «сначала я искал глаза, а потом понял, где должна быть остальная часть лица». Тут только говорится, к какому выводу она пришла. Нужны алгоритмы, которые выводят: «Я нахожу новый шаблон в большинстве моих данных, давайте дадим ему имя».

Насколько трудно подняться на этот уровень скажут программисты. Пока же этого уровня нет. Так, изменение изображения незаметным для человека способом может привести к тому, что нейронная сеть назовет изображение другим понятием.

## **Головной мозг человека как основа конструирования искусственного интеллекта**

Теория глобального нейронного рабочего пространства (GNWT), одна из доминирующих научных теорий сознания. Теория начинается с мозга и делает вывод, что некоторые из его особенных архитектурных особенностей порождают сознание. Её происхождение можно проследить до «архитектуры классной доски» компьютерной науки 70-х годов XX века, в которой специализированные программы обращались к общему хранилищу информации, называемому классной доской или центральным рабочим пространством.

Такой ресурс обработки существует в головном мозге человека и является центральным для человеческого познания. Его ёмкость мала, поэтому только один образ, мысль или воспоминание занимает рабочее пространство в любой момент времени. Новая информация конкурирует со старой и вытесняет её. Деэн и Шанжё постулировали, что рабочее пространство представлено сетью пирамидных возбуждающих нейронов, связанных с кортикальными областями, в частности, префронтальными, парието темпоральными и срединными поясными ассоциативными областями. Активность головного мозга человека остаётся локализованной и поэтому бессознательной, например, активность модуля, контролирующего, куда смотрят глаза, или модуля, регулирующего положение тела.

Но когда активность в одном или нескольких регионах превышает порог, это запускает волну нейронного возбуждения, которая распространяется по всему нейронному рабочему пространству, по всему головному мозгу человека. Эта сигнализация становится доступной для множества вспомогательных процессов, таких как язык, планирование, схемы вознаграждения, доступ к долговременной памяти и хранение в буфере кратковременной памяти. Акт глобального распространения этой информации делает её сознательной. Осознанный опыт формируется пирамидными нейронами, контак-

тирующими с областью моторного планирования головного мозга человека. Другие модули передают сообщение об ожидании эффекта.

Сознательные состояния возникают из того, как алгоритм рабочего пространства обрабатывает соответствующие сенсорные входы, моторные выходы и внутренние переменные, связанные с памятью, мотивацией и ожиданиями. Глобальная обработка является сознанием.

Нейронную сеть не учат «понимать». Во время фазы обучения она корректирует внутренние связи в симулированных нейронных сетях, чтобы наилучшим образом предугадать следующее слово, слово после него и так далее. Обученная на бесчисленных веб-страницах, блогах, рекомендациях и книгах, она содержит до тысячи миллиардов соединений, эмулирующих синапсы, т.е. соединения, где один нейрон взаимодействует со следующим. Такая же схема неконтролируемого обучения предсказанию следующего токена применяется к программному коду, уравнениям и логическим выражениям, таким как в математике или теоретической физике.

Интегрированная информация является математической мерой, количественно определяющей, сколько «внутренней причинной силы» обладает некоторый механизм. Нейроны, запускающие потенциалы действия, которые влияют на клетки ниже по потоку, к которым они подключены через синапсы, являются одним типом причинного механизма, как и электронные схемы, состоящие из транзисторов, емкостей, сопротивлений и проводов.

Внутренняя причинная сила может быть точно оценена для любой системы. Чем больше её текущее состояние определяет причину (вход) и свой эффект (выход), тем больше причинной силы она обладает.

Теория глобального информационного пространства постулирует, что любой механизм с внутренней силой, чьё состояние нагружено его прошлым и его будущим, обладает сознанием. Важно иметь полную спецификацию точной синаптической проводки всей нервной системы.

Если для создания сознания достаточно имитировать функциональность мозга, как постулирует теория GNWT, симулированный человек будет

обладать сознанием, перевоплотившись внутри компьютера. Симулякр будет чувствовать столько же, сколько программное обеспечение. Он будет действовать как человек, но без каких-либо врожденных чувств.

Если компьютеры ощущают жизнь через свои собственные чувства, они перестают быть средством для достижения цели, определяемой их полезностью для людей. Они становятся целью сами по себе. Согласно GNWT, они превращаются из простых объектов в субъекты. Каждый существует как «Я» с точкой зрения.

Теория квантового сознания утверждает, что мысль возникает в результате взаимодействия частиц на субатомном уровне. Эта идея была выдвинута еще в 80-е годы XX века, но ее недостатком было почти полное отсутствие понимания того, как это может происходить. Масштабно организованные события на квантовом уровне всегда наблюдались только в лабораторных условиях при тщательном исключении температурных и других хаотических эффектов, что никак не сопоставимо с теплой и влажной субстанцией головного мозга человека.

Анестезиолог Стюарт Гамеров предположил, что местом воплощения квантового сознания могут быть микротрубки. Это цилиндрические молекулы диаметром примерно 25 нм. Микротрубки сопрягаются друг с другом как элементы конструктора, образуя внутренний скелет клетки. Это считалось их главной функцией. Однако они обладают способностью переходить из одного состояния в другое за наносекунды и могут легко кооперироваться, обрабатывая информацию в одноклеточных организмах. Пока не известен механизм, который позволял бы им взаимодействовать с другими микротрубками через стенки клетки.

### **Опасности применения искусственного интеллекта**

Существуют два вида опасности применения ИИ. Это ошибки программного обеспечения и самостоятельная активность искусственного интеллекта во вред человеку. Опасность первого рода характерна для всякого

программного обеспечения. Важно определить, какая степень интеллекта может быть допустимой для ИИ. Сближение интеллектуальных возможностей ИИ и человека вынуждает точно измерять эти свойства.

Интеллектуал, будучи человеком, обладает профессиональными компетенциями решения задач. Продуктом применения его интеллекта является новая теория (ученый), прикладное применение существующих теорий (инженер и менеджер). Умных людей больше, чем интеллектуалов.

Нужно различать психические и логические способности к решению задач. Интеллектуал применяет знания, умный человек создает знания. ИИ может содержать логические ошибки. Поскольку решаемые им задачи могут быть очень сложными, а список задач заранее не известен, необходимо предусматривать специальные меры для верификации полученного решения.

Интеллектуальные системы основаны на знаниях человечества. Поэтому наделение ИИ способностью людей делать выводы о ментальном состоянии других людей является ключевой задачей. Это позволит ИИ интегрироваться в общество.

На уровне технологических решений это означает, что беспилотные автомобили будут понимать и делать выводы о психическом состоянии водителей и пешеходов. Это поможет ИИ предсказывать поведение и снижать количество автомобильных и воздушных аварий.

В робототехнике используется процесс обучения на основе технологии распознавания образов и знаний из моделей для развития интеллекта со здравым смыслом. Труднее передать искусственному интеллекту механизмы социального взаимодействия и морали человека.

Опасения связаны с потенциалом искусственного интеллекта, основанного на теории разума, предсказывать и понимать поведение и намерения человека. Это может привести к навязчивому наблюдению или манипуляциям.

## **Прикладные перспективы теории разума искусственного интеллекта**

Технологии ИИ функционируют в настоящий момент на трёх уровнях их возможностей. Реактивный ИИ не имеет функции памяти и прогнозирует результаты на основе полученных входных данных. На уровне ограниченной памяти ИИ использует эту память для обучения и улучшения диалоговых реакций. Теория разума реализует третий уровень ИИ.

На этом уровне технология ИИ реализует принцип обратной связи с пользователем сети. Эта функция ИИ воспринимается как понимание потребности других разумных существ.

Технологии ИИ, базирующиеся на теории разума, находятся на этапе разработок. Целью разработок является наделение технологии ИИ функцией обратной связи. Она является важным условием прикладного применения теории разума. Нейронная сеть, основанная на теории разума, имеет механизм глубокого машинного обучения.

Теория разума ИИ будет реализована с помощью системы машинного обучения, которая может объяснять решения на разных языках, помогая пользователю (человеку) понимать их. Нил Рабинович и его команда создали систему искусственного интеллекта на основе теории разума. Эта система наблюдает за другими системами искусственного интеллекта и изучает их характеристики и функции. Она состоит из трёх искусственных нейронных сетей. Это попытка имитировать сеть нейронов, из которых состоит человеческий мозг. Имитация нейронных сетей позволяет компьютерам принимать решения так же, как и человеку.

Система состоит из небольших вычислительных элементов и связей, которые обучаются на основе опыта, стремясь походить на человеческий мозг. Первая искусственная нейронная сеть изучает тенденции других систем на основе их прошлых действий. Вторая искусственная нейронная сеть формирует более глубокое понимание убеждений текущей системы искусствен-

ного интеллекта. Третья искусственная нейронная сеть использует выходные данные двух других систем и прогнозирует следующие действия искусственного интеллекта.

Система на основе искусственного интеллекта «Теория разума» понимает, какие вербальные и невербальные сигналы следует учитывать. Сложно дифференцировать их на основе таких элементов, как эмоции, поскольку существуют разные уровни эмоциональной зрелости независимо от возраста. Чтобы понять, почему люди реагируют именно так, а не иначе, необходимо учитывать различные факторы.

Технология «Теории разума» предполагает создание ментальных моделей других разумных существ, которыми на данном этапе являются люди.

Искусственный интеллект сосредоточен на понимании и прогнозировании ментальных состояний других людей, что не является прямой задачей других технологий искусственного интеллекта.

Генеративный ИИ создает новый контент на основе изученных шаблонов, тогда как обучение с подкреплением оптимизирует процесс принятия решений с помощью вознаграждений. Способность ИИ, основанная на теории разума, позволяет ему строить более тонкие социальные взаимодействия. Генеративный ИИ создаёт контент целенаправленного поведения, характерного для обучения с подкреплением. Но ИИ сталкивается с ограничениями в точной интерпретации сложных человеческих эмоций и намерений, что менее выражено в других областях ИИ.

К информационно-технологическим направлениям развития автоматизированных систем относят инжиниринг искусственного интеллекта (AI engineering), основанный на применении методологии автоматизации технологических процессов сборки, настройки и развёртывания программного обеспечения – DevOps, процессно-ориентированной методологии разработки и предоставления аналитических данных – DataOps, методологии эффективного развёртывания и поддержки моделей машинного обучения на производстве – MLOps; доверенный искусственный интеллект (Responsible AI); ком-

позитный искусственный интеллект (Composite AI); искусственный интеллект для промышленного Интернета вещей (Edge AI); интеллектуальные приложения и помощники; облачные сервисы и автономные системы.

Компании создают агентский режим внутри собственной экосистемы ИИ-ассистентов, которая охватывает жизненный цикл разработки от сбора требований до тестирования и сопровождения готового решения. В клиентском режиме ИИ-ассистент выполняет последовательность действий по заданию инженера. В отличие от классических подсказчиков кода агентский режим превращает инструмент в полноценного «цифрового коллегу». Он понимает задачу, способен выполнять сценарии, работать с контекстом проекта, учитывая структуру репозитория, может пользоваться утилитами, командами и создавать файлы. ИИ-агент доступен в редакторе, выполняет инструкции на русском и любом другом языке.

Достоинством инструмента является интеграция с существующей инфраструктурой и понимание контекста больших проектов экосистемы. Агентский режим ускоряет разработку и ревью, объединяя десятки действий в один запрос. Он помогает быстрее выполнять стандартные операции, упрощает старт новичкам и повышает продуктивность опытных разработчиков, оставляя больше внимания архитектуре и продукту. Время на повторяющиеся операции сокращается.

Экосистема ИИ включает спектр ИИ – инструментов для анализа, разработки, тестирования, внедрения и последующего сопровождения продукта. AI-search помогает ИТ-командам быстро находить информацию, способы починки бага, бизнес-правила и архитектурные решения.

AI-Data (AI-помощник) помогает писать и править SQL-запросы, генерирует релевантные скрипты для ETL и отвечает на аналитические вопросы в контексте работы с данными. Autodesc генерирует релевантные описания таблиц и полей и публикует их в дата – каталоге. T-Cover Agent (LLM-агент) генерирует корректные unit-тесты, повышает тестовое покрытие и защищает логику от регрессий.

T-Weaver анализирует документацию и генерирует полный набор тестовых сценариев со структурой, шагами, ожидаемыми результатами и нужными вложениями. T-Code Review (AI-помощник для автоматизации кодревью) покрывает 100% всех репозиторий. Это 1 000+ комментариев в день.

AI-ассистент (умный AI-ассистент) используется для разработки в IDE. Safeliner (AI-ассистент) используется для задач информационной безопасности. SRE-ассистент (цифровой SRE-инженер) ищет информацию и документирует её. Anomaly Analyzer в режиме реального времени обнаруживает отклонения в метриках и присылает командам уведомления.

## **Заключение**

Проведённый анализ особенностей эволюции технологий искусственного интеллекта показал, что теория разума указывает на третий этап совершенствования интеллекта на технологической основе программных средств и инструментов технической кибернетики. Под искусственным интеллектом понимается формальное мышление вычислений, которым разработчики наделили компьютерные программы.

Операционную основу этого интеллекта формирует математическая логика. С помощью её логических средств созданы вычислительные мощности, которые имеют аппаратную часть. В функционировании этих мощностей важную роль играют инструменты и модели.

Математический формальный интеллект вычислений основан на инструкциях и программах действий, которые закладываются в технологию через код и реализуются на основе алгоритмов решения конкретных задач деятельности, связанных с управлением, сбором и обработкой данных, диалогом с пользователями, автоматизацией и созданием экосистем умных городов, виртуальных предприятий и виртуальных электростанций.

Вычислительная формальная функция интеллектуального мышления не тождественна всему разнообразию функций мышления человека. В этой функции нет ценностной тематики и эмпатии.

Эта функция является инструментальной. Она ограничена синтаксическими задачами формализации информации для трансформации её в различные способы трансляции данных. В сети информация передаётся в виде сигнала. На входе она кодируется в форму сигнала через использование двоичного кода. На выходе она декодируется в формат цифрового двоичного кода, структуру которого формируют операции формальной логики, которые имеют модификацию математической логики.

Вычислительный интеллект первоначально был ограничен синтаксисом и формальной обработкой информации и данных на основе закрытых алгоритмов. Этот искусственный интеллект называют реактивным. На следующем этапе эволюции вычислительного искусственного интеллекта стали использоваться алгоритмы открытого типа с ограниченной памятью, которые стали сопрягаться с машинным обучением и глубоким машинным обучением. В результате вычислительный интеллект стал оперировать операционными структурами нечётких логик. Вследствие этого возникли предпосылки для создания диалоговых компьютерных программ, использующих алгоритмы, основанные на логике вопросов и ответов.

Но чат боты так и не приобрели функции, свойственные интеллекту человека. Позиции сохранил формализм подбора нужного ответа из заданной базы ответов под заданный вопрос. Этот подбор осуществляется посредством вычисления правильного ответа. Но в этой модели искусственного интеллекта с ограниченной памятью нет эмпатии. Алгоритмы вычислений не предполагают эмпатии. Они не могут её симулировать. А это снижает эффект общения с искусственным интеллектом. Разработчики обнаружили, что люди склонны к антропологизации даже искусственного интеллекта. Но симуляция искусственного интеллекта с ограниченной памятью слишком неубедитель-

ная и не способствует восприятию этого интеллекта как автономного существа по аналогии с другим человеком.

И тогда разработчики обратились к теории разума искусственного интеллекта, поскольку она предполагает реализацию более высокого уровня симуляции понимания особенностей интеллекта другого человека, и другого агента искусственного интеллекта. Это уровень вычислительного семантического искусственного интеллекта. Этот уровень предполагает использование эмоционального дизайна в разработке интерфейсов. Толчок дала гейм – индустрия, которая разработала системную методологию использования предметной онтологии, компьютерного зрения и распознавания образов.

В результате вычислительный искусственный интеллект благодаря разработчикам приблизился к освоению фундаментального принципа кибернетики. Это принцип обратной связи. Он предполагает, что взаимодействия играют ключевую роль в эмансипации агентов экосистемы искусственного интеллекта. Важным шагом в эволюции технологий вычислительного искусственного интеллекта стала разработка нейронных сетей, которые имитируют не только мышление человека, но и отдельные семантики его сознания. Такими семантиками стали автономность мышления, самообучение и формирование ресурсов оперативной и долгосрочной памяти в виде библиотек, блокнотов и методов. На этой основе нейронные сети приобрели функцию писать код, т.е заниматься программированием.

Но по мере эволюции вычислительного искусственного интеллекта стали проявляться риски, связанные с предубеждением и галлюцинациями. Причиной является недостаток вычислительных мощностей и ограниченность баз данных. В результате вычислительный искусственный интеллект может импровизировать и генерировать не свойственные людям варианты решений поставленных ему задач.

Одной из проблем является качество данных. Оказалось, что в базах данных присутствуют не только данные, но и их интерпретации в редакциях

конкретных людей со свойственной им предвзятостью и галлюцинациями. Инженерам в области софта приходится решать эту задачу.

## ЛИТЕРАТУРА

1. Аджемоглу, Д., Робинсон, Дж. А. Почему одни страны богатые, а другие бедные: происхождение власти, процветания и нищеты / пер. с англ. Д. Литвинова, П. Миронова, С. Сановича. М.: АСТ., 2015- 692 с.

2. Loiko, A. I. Big Data Ethics: a textbook on the academic discipline "Philosophy and Methodology of Science" for master's students of all specialties / A. I. Loiko; Belarusian National Technical University, Department of «Philosophical Teachings». – Minsk: BNTU, 2025. – Деп. в БНТУ 15.05.2025, № DEPBNTU-2025-244.

3. Loiko, A. I. Big Data Ethics: a textbook on the academic discipline "Philosophy and Methodology of Science" for master's students of all specialties / A. I. Loiko; Belarusian National Technical University, Department of «Philosophical Teachings». – Minsk: BNTU, 2025. – Деп. в БНТУ 15.05.2025, № DEPBNTU-2025-244.

4. Cognitive artificial intelligence: philosophy digital law and ethics: Textbook on general education discipline "Philosophy and Methodology of Science" for advanced higher education / A. I. Loiko; Belarusian National Technical University, Department of «Philosophical Teachings». – Minsk: BNTU, 2024. – Деп. в БНТУ 20.11.2024, № DEPBNTU-2024-222.

5. Philosophy of artificial intelligence [Electronic Resource]: textbook on general education discipline "Philosophy" for students of all forms of study / A. I. Loiko; Belarusian National Technical University, Department of «Philosophical Teachings». – Minsk: BNTU, 2024. – Деп. в БНТУ 17.09.2023, № DEPBNTU-2024-213.

## ОГЛАВЛЕНИЕ

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| ВВЕДЕНИЕ.....                                                                        | 3  |
| Теория разума искусственного интеллекта агентов.....                                 | 3  |
| Экосистема агентов искусственного интеллекта.....                                    | 8  |
| Этические аспекты теории разума искусственного интеллекта .....                      | 20 |
| Когнитивные функции искусственного интеллекта – агента.....                          | 21 |
| Понятие интеллекта .....                                                             | 36 |
| Теория множественного интеллекта .....                                               | 40 |
| Интеллект и сознание .....                                                           | 43 |
| Искусственный интеллект .....                                                        | 52 |
| Вычисление .....                                                                     | 54 |
| Распределённые вычисления .....                                                      | 57 |
| Когнитивные вычисления .....                                                         | 58 |
| Модификации вычислительного интеллекта .....                                         | 60 |
| Вычислительный интеллект и мягкие вычисления .....                                   | 61 |
| Вычислительное мышление и программирование .....                                     | 68 |
| Искусственный интеллект и феномен сознания .....                                     | 70 |
| Коннекционизм и нейронные сети .....                                                 | 72 |
| Искусственные квалиа .....                                                           | 76 |
| Проблемы искусственного интеллекта .....                                             | 77 |
| Сознание и большие данные .....                                                      | 80 |
| На пути к искусственному интеллекту .....                                            | 82 |
| Головной мозг человека как основа конструирования искусственного<br>интеллекта ..... | 83 |
| Опасности применения искусственного интеллекта .....                                 | 85 |
| Прикладные перспективы теории разума искусственного интеллекта.....                  | 87 |
| Заключение .....                                                                     | 90 |
| ЛИТЕРАТУРА .....                                                                     | 93 |