Что мы знаем об интеллекте?

Интеллект не универсален и определяется различными факторами: хардвером, софтвером, ветвером (wetware из научной фантастики), гормональной регуляцией, средой, опытом, знаниями, контекстом и т.д.
Существует многомерный спектр интеллекта, точные измерения которого пока что недоопределены, однако в эмпирическом опыте мы видим, что существуют кластеры в этом многомерном спектре, разметка которых осуществляется из разных онтологий. Вот пара примеров, которые относительно хорошо исследованы: пространственный-лингвистический спектр интеллекта или аналитико-эмоциональный спектр.
Даже внутри человеческой популяции наблюдается не единая норма интеллекта, а варианты нормы со своими особенностями конфигурации, субъективно воспринимаемыми как плюсы и минусы: развитость в пространственном мышлении ассоциирована с повышенным риском аутизма, а на другой стороне этого спектра - развитость лингвистического мышления ассоциирована с повышенным риском шизофрении.

Понятно, что для человеческого интеллекта правильная идеализация эталона - это не точка, а, по видимости, странный аттрактор в многомерном пространстве спектров интеллекта.
Как и со всеми сложными признаками, профиль индивида в пространстве типов интеллекта определяется отчасти генетикой, отчасти средой. В какой степени, как и какие гены и этапы феногенеза определяют те или иные типы мышления - мы в точности не знаем. Но в любом случае: опыт, априорные знания, когнитивные метафоры, сформировавшиеся в ходе развития индивида - оказывают колоссальное влияние не только на несуществующую в объективных метриках интегральную силу интеллекта, но и профиль развития интеллекта в многомерном пространстве.

Помимо этого, крайне важны внешние по отношению к индивиду феномены: эмерджентность роевого интеллекта, экзокортекс и внешние артефакты-усилители интеллекта (от письменности до яндекс-навигатора), использование естественных феноменов как части вычислительной схемы

(магнитный север как часть системы пространственного интеллекта перелетных птиц).
Внешние системы могут радикально изменить эффективность и профилизацию интеллекта.

Все эти свидетельства не оставляют никаких оснований полагать, что машинный интеллект в принципе будет таким, как интеллект человека, даже в случае если мы потратим неоправданно большие усилия на полное воспроизводство внутренних и внешних условий. Возможно ли полное эмулирование и повторение человеческого интеллекта в компьютере - вопрос интересный научно, но в практической, экономически обоснованной плоскости вопрос не стоит в обозримой перспективе.

Машинный интеллект и открытые вопросы

Машинный интеллект будет иметь другой профиль развития. Какой именно - будет зависеть от того, каким мы создадим ИИ: от архитектуры процессоров (графический процессор хорош для решения одних задач, центральный процессор - других, и нет оснований полагать, что даже нейроморфный процессор будет полностью совпадать по параметрам с живым нейроном), программной архитектуры, методам обучения, парадигме программирования.

Машина точно сильнее человека в вычислениях, решении сравнительно узких задач, обработке больших данных, многократно превосходит по размеру оперативную память человека.

Мы не знаем, в каких других задачах удастся создать эффективную конкуренцию человеку. Не знаем, создадим ли мы одну универсальную (AGI) или несколько разных систем узкого искусственного интеллекта, каждая из которых будет хороша в своем классе задач.

Не знаем, сможет ли машина, как человек, учиться не по миллиарду, а по одному датапоинту, и глубоко присваивать результаты такого обучения, как на это способен человек (психологическая травма, ПТСР - предельный

случай, когда один эпизод многократно переживается и рефлексируется; результат этой рефлексии становится ядерной частью контура принятия решений человека).

Не знаем, будет ли ИИ обладать сверхчеловеческим контрфактуальным мышлением, воображением, способностью строить мыслительные конструкции "что, если мир был бы устроен не так, а иначе".

В конце концов, мы не знаем, является ли человеческий нейрон классическим или квантовым вычислителем.

Взаимодействие и конкуренция, проблема сильного ИИ

Вопрос доминирующий в общем дискурсе (особенно среди философов и политиков, а не ученых и инженеров) "станет ли ИИ сильнее человека?" представляется неинтересным и непродуктивным.

Интересными теоретически, но операционно преждевременными представляются вопросы:
1. По каким типам интеллекта целесообразно (то есть в первую очередь - экономически эффективно) будет использовать ИИ, а не человека: целеполагание, стратегическое мышление, контекстночувствительное мышление, межпредметное мышление - главные кандидаты.
2. Как осуществляется QA (quality assurance, контроль качества) систем ИИ? Как построен explainable AI, как решать проблему черного ящика?
3. Какие практические задачи мы на самом деле хотим делегировать ИИ, а какие - даже отбросив предрассудки и романтические представления о антропоцентричности мира - нет.
4. По какой схеме складывается экономически обоснованная реальность по основной массе задач, которые мы не готовы делегировать ИИ:
- Human in the loop - человек в цепочке принятия решения. Интеллектуальный профиль человека и машины формируются так, что человека и машина эффективно дополняют друг друга по этим задачам.

- Human on the loop - человек над цепочкой принятия решения. Участие человека целесообразно не в проработке вариантов решения, но в контроле качества, выборе решения и верификации результатов работы ИИ.
- Human out of the loop - человек вне цепочки принятия решения.
5. Как устроено общество, социальные институты, управление политикой и общественный контроль в эпоху цифровых кентавров.

Программа исследований и разработок

Специфика ситуации такова, что программу приходится формировать в ситуации, когда понимания принципиальных возможностей и профиля интеллекта будущих ИИ у нас нет и в короткой перспективе не будет.
Это означает, что необходимо определить направления исследования, важные инвариантно от того, какими будут ответы на ключевые вопросы. Если эта задача будет решена успешно, то вне зависимости от пути, по которому пойдет ИИ, результаты работы проекта "Кентавр" смогут быть использованы для формирования "практик будущего" в понимании Кружкового движения - то есть сформированных ячеек, которые сегодня начинают проращивать те практики, которые впоследствии станут нормой общества эпохи кентавров.
Гипотезы, с которыми работает проект кентавр, антихрупкие по Талебу - даже в ситуации невероятного сценария полного превосходства ИИ над человеком по важным для нас задачам, набор сформированных практик будет тем более полезным:
1. Интерфейс взаимодействия человека и машины является точкой для эффективного и быстрого возврата инвестиций в развитие системы по мере достижения приемлимого уровня уровня развития ИИ и человеческого капитала.
Классический пример кентавров Каспарова - команда шахматистов-любителей с тремя ноутбуками обыгрывает гроссмейстеров с суперкомпьютерами, так как придумывает, как правильно построить работу с машиной.
Инвестировать в интерфейс кентавра - выгоднее всего, и текущего уровня нейросетей даже в задачах с на первый взгляд низкой эффективностью ИИ

достаточно, чтобы наибольшего результата можно было достичь через правильный интерфейс взаимодействия.
2. Ключевая компетенция человека эпохи кентавров - дизайн кентавров.
Хороший дизайн кентавра - открытая система с положительными обратными связями между человеком и ИИ, которая обеспечивает самоулучшение как всей системы, так и отдельных элементов.
Умение анализировать ситуацию и делать правильные гипотезы о том, как хороший интерфейс построить - компетенция, которую необходимо нарабатывать практикой. Положительные обратные связи снежным комом выведут систему на недостижимый уровень эффективности независимо от индивидуальных возможностей человека и ИИ.

Универсальная формула, обеспечивающая устойчивость дальнейшего развития в любом из возможных сценариев:
Человек + ИИ > ИИ

Для того, чтобы научиться создавать человек-машинные системы (кентавры) по этой формуле, необходимыми представляются два направления работ:

1. **Создание универсальной методологии кентавров**: набора принципов, по которым строятся эффективные связки из людей и ИИ. Для создания такой методологии подходят задачи, в которых ИИ уже показывает недостижимый уровень, например шахматы и го. Эффективно созданные кентавры, решающие смежные задачи: обучение игре в го, усиление человеческого интеллекта, развитие метапредметных компетенций человека через взаимодействия в составе кентавра - позволят нащупать универсальные принципы организации взаимодействия, которые в обобщенном виде в дальнейшем могут быть применены к тем областям, где конкурентный ИИ только создается.
2. **Создание систем взаимного обучения человека и ИИ**, в которых благодаря методологии и дизайну кентавра создаются цепочки положительных обратных связей как для человека, так и для ИИ. И человек, и ИИ дообучаются в ходе эксплуатации кентавра и развивают свои сильные компетенции. Более того, ключевое

требование кентавров по этому типу разработок является решение экономически востребованных задач с перспективой взрывного роста - только взрывной рост, обеспеченный рыночной тягой, позволит достаточно быстро масштабировать практики проектирования кентавров.

В первой волне в программе “Кентавр” (2020-2021) большую часть занимают задачи первого типа. По мере накопления знания о кентаврах, со второй половины 2021 года, все большую часть занимают задачи по созданию кентавров в рыночно перспективных задачах.

Критерии отбора кейсов первого типа:

1. Методическая чистота эксперимента. ИИ сильнее человека уже сегодня по всей или части задачи - для того, чтобы убедиться, что методологические выводы исследования верны по отношению к ИИ, более сильному, чем человек.
2. Высокий уровень научного результата. Обеспечивается отбором кейсов со статусными партнерами (мировыми экспертами в своем поле), верифицирующими результат.
3. Быстрая обратная связь. Полезные выводы достижимы в коротком горизонте, после 2-5 месяцев разработки.

Результат упаковывается в прототипы сервисов, отдается в OpenSource, выводы публикуются.

Пример: игра в го.

Человек уже не может обыграть компьютер, однако вполне возможно решение задач по повышению уровня человека (человек с уровнем 5 дан + ИИ с уровнем 5 дан обыгрывают ИИ с уровнем 6 дан), более эффективному обучению человека игре в го, повышение уровня присвоения и переноса в другие области компетенций, формирующихся при игре в го: стратегическое мышление, видение общей картинки, тактическое мышление и др.

Критерии отбора кейсов второго типа:

1. Уже есть слабый, узкий ИИ. Для кейса существует ИИ, который уже решает узкую задачу на сопоставимом с человеком уровне, но не может обеспечить решения широкой задачи.
2. Market pull. Успешное решение обладает коммерческим потенциалом для быстрого развития в логике НТИ.
3. Есть гипотезы о том, как эффективно сформировать взаимнообучающуюся связку из человека и ИИ с петлями положительной обратной связи, приводящую к развитию интеллекта человека и машины в разных направлениях.
4. По результатам отрботки кейса формируются ячейки “Практик будущего”, которые используют кентавра в своей деятельности, повышают конкурентноспособность и обучают сообщества новым принципам работы.

Результат упаковывается в коммерчески устойчивые, растущие бизнесы, сообщества проектировщиков и пользователей кентавров.

Пример: лечение рака.

Онколог средней квалификации уже сегодня дает результат по диагностике сопоставимый с нейросетью.

Кентавр с хорошим дизайном обеспечивает:

- дообучение нейросети на локальных данных, в том числе персонализацию под условия конкретного врача, - например, через дообучение с учетом особенностей спектра местных фенотипов, образа жизни и прочих условий.

- инструментарий и средства для повышения квалификации и развития в других направлениях онколога; врач начинает смотреть не на снимок, а на пациента, его анмнез, внешний вид, учитывает его особенности и постепенно становится специалистом, видящим широкую картину, которую эта нейросеть не может анализировать.

- сам интерфейс взаимодействия человек-машина гибок, интерактивен и дорабатывается под сценарии использования, персональную траекторию врача и условия работы.