

Анатолий Левенчук



ИНТЕЛЛЕКТ-СТЕК
2023

Учебник

Анатолий Левенчук
Интеллект-стек 2023

«Издательские решения»

Левенчук А.

Интеллект-стек 2023 / А. Левенчук — «Издательские решения»,

ISBN 978-5-00-604990-1

Курс «Интеллект-стек» раскрывает понятия интеллекта и мышления, предлагает усиление естественного и искусственного интеллекта за счёт изучения лучших (SoTA, state-of-the-art) мыслительных практик, основанных на трансдисциплинах понятизации, собранности, семантики, математики, физики, теории понятий, онтологии, алгоритмики, логики, рациональности, познания/исследований, эстетики, этики, риторики, методологии, системной инженерии.

ISBN 978-5-00-604990-1

© Левенчук А.
© Издательские решения

Содержание

Введение	6
1. Интеллект и практики интеллект-стека	8
Берём понятие интеллекта у разработчиков машинного интеллекта	8
Интеллект – это вычислитель,	10
Нет наград, но есть избегание неизвестности. Оптимизм	14
Интеллект – это вычислитель	27
Кейс по различению интеллекта и мышления	31
Прикладное мастерство,	32
Интеллект и прикладное мастерство неразрывны, мышление и прикладное рассуждение тесто переплетены	35
Бытовое использование терминов, которые мы использовали при рассказе об интеллекте и мышлении	40
Бесконечное развитие требует интеллекта	44
Интеллект определяет скорость обучения новому мастерству	48
Интеллект врождённый и приобретённый	51
Трансдисциплинарный интеллект-стек	55
Без владения трансдисциплинами интеллект-стека трудно понимать тексты, сосредоточиться, рассуждать	59
Граница между мыслительным и прикладным мастерством размыта	62
Важен порядок решения проблем:	64
Кейс: интеллект как способность научиться	67
Мышление – это функция/поведение/назначение интеллекта	69
Чему учиться уже образованным?	72
Упражнение:	80
Интеллект против культа карго	82
Дикарский мир 21 века: переучиваться самому, жалеть и учить других	84
Трансдисциплины: название то же, содержание уже другое	86
2. Понятизация	92
Быстрое мышление через поиск ассоциаций	92
Передача субъективного опыта	95
Нейросемиотика. Нейросемиотическое программирование	101
Психопрактики для понятизации: внимание к кинестетике	107
3. Собранность	109
Агенты	109
Собранность: владеть умом и телом	114
Собранность как часть актёрского/ролевого мастерства	118
Конец ознакомительного фрагмента.	119

Интеллект-стек 2023

Анатолий Левенчук

© Анатолий Левенчук, 2023

ISBN 978-5-0060-4990-1

Создано в интеллектуальной издательской системе Ridero



Введение

Курс «Интеллект-стек» раскрывает понятия интеллекта и мышления, а затем предлагает усиление естественного и искусственного интеллекта за счёт изучения лучших (SoTA, state-of-the-art) мыслительных практик, основанных на трансдисциплинах понятизации, собранности, семантики, математики, физики, теории понятий, онтологии, алгоритмики, логики, рациональности, познания/исследований, эстетики, этики, риторики, методологии, системной инженерии. Курс входит в пятый семестр программы «Организационное развитие» Школы системного менеджмента.

В курсе говорится о понятии «интеллект» и его поведении «мышление», вводится понятие интеллект-стека как набора мыслительных практик (методов мышления, паттернов мыслительной деятельности), позволяющих интеллекту быстрее решать всё более широкий спектр проблем, мешающих изменить себя и мир к лучшему. Улучшение мира понимается как многоуровневое (например: вещество, существо (включая людей), популяция, вся жизнь на Земле) избегание неприятного сюрприза в будущем, в пределе – снижение экзистенциальных рисков¹, «спасение». Сам состав практик стека и последовательность изложения материала курса специально адаптированы для учебных целей.

Для каждой мыслительной практики интеллект-стека говорится о том, зачем она нужна, какое её место в мышлении, а также приводится краткая характеристика её современного (SoTA) состояния, даются ссылки на литературу. Содержание курса предполагает пререквизитом прохождение курсов первых четырёх семестров программы «Организационное развитие» Школы системного менеджмента. Курс не пересказывает материал трансдисциплин, изучаемый в других курсах, но кратко касается содержания каждой из трансдисциплин интеллект-стека. Курс предназначен для того, чтобы как-то «склеить» вроде бы разрозненные знания трансдисциплин из многих других курсов в целостную картину мира. Основная идея в том, что это «аспирантский» курс: он должен вывести понимание текущей ситуации в познании на фронт, подготовить к самостоятельным исследованиям, к самостоятельному планированию своего развития в контексте новейших достижений цивилизации. Курс сам по себе не столько «учебный курс», сколько аннотированный обзор литературы: основные знания содержатся в литературе, на которую ссылается курс, а материал курса только указывает на то, что именно надо вычитывать в указанной литературе. В курсе более семисот ссылок на литературные источники (хотя небольшая часть из них дублируется из соображений удобства использования).

Текущая версия интеллект-стека опирается на исследования учёных, занимающихся проблемами наиболее общих закономерностей эволюции жизни (дарвиновской эволюции) и эволюции знания и инженерных систем (техно-эволюции). Это труды групп David Deutsch, Виталия Ванчурина, Karl Friston, Андрея Хренникова, Kit Fine, John Doyle, Michael Levin, Andrej Karpathy и многих других учёных, с которыми они сотрудничают. Основная идея – это базирование объяснений интеллекта на первых принципах, выводимых из физики (термодинамика) и теории эволюции. Если очень кратко сформулировать основные идеи получившейся компиляции современного состояния трансдисциплин, то это будет неумолимое увеличение сложности многоуровневых биологических и технических систем в ходе биологической и техно-эволюции, безмасштабность и неантропоцентричность объяснений, творчество в основе деятельности, нацеленность агентов (естественных и искусственных, а также коллективных) на изменение мира к лучшему как пути «спасения».

¹ https://en.wikipedia.org/wiki/Global_catastrophic_risk

В курсе в текущей версии приведено не так много заданий. Возможно, в следующих версиях курса число заданий будет увеличено. В любом случае, к моменту прохождения этого курса предполагается, что студент достаточно организован, чтобы самостоятельно спланировать своё обучение. Пока не планируется иметь варианты прохождения этого курса с инструктором или преподавателем, курс нацелен на самостоятельную проработку материала в удобном для студента темпе. Основное время должно уходить тут не столько на освоение самого материала курса, сколько на знакомство с указанной в курсе литературой.

После прохождения курса студент должен демонстрировать понимание того, что такое интеллект и мышление, каким образом их можно усилить. Студент должен демонстрировать целостное мировоззрение, позволяющее связно и неантропоцентрично описывать многомасштабный мир, строить свои R&D программы (инженерные программы, включая их часть, связанную с исследованиями). После прохождения курса ожидается, что студент организует свою лабораторию, которая в ходе шестого семестра программы «Организационное развитие» разработает и будет развивать курс по какой-то прикладной практике инженерии (включая «железную» инженерию, менеджмент, образование, медицину и т.д.). Это ход на подъём квалификации со студенческого «мастера» до аспирантского «реформатора», то есть переход с ведения программы организационного развития какого-то предприятия на программу развития сообщества какой-то практики.

Курс получен путём полной переписки содержания курса «Образование для образованных 2021». Содержание курса и терминология согласованы с содержанием и терминологией курсов программы «Организационное развитие» ШСМ. Ситуация с лучшим на сегодня известным знанием и пониманием того, что такое интеллект, быстро меняется. Обновления курса будут частыми и после окончания переписки варианта 2021 года.

Благодарность студентам и сотрудникам Школы системного менеджмента, которые поддерживали меня в ходе этой работы. Особая благодарность Роману Варьянко, который оперативно выполнил корректуру текста переписанных разделов, не ограничиваясь грамматикой и орфографией, но и делая содержательные замечания.

Ваши вопросы и замечания по текущей версии, предложения по поводу следующих версий курса давайте в чат поддержки, он организован в телеграм как общий для всех курсов пятого семестра «Образование для образованных» программы «Организационное развитие»: https://t.me/odo_course

1. Интеллект и практики интеллект-стека

Берём понятие интеллекта у разработчиков машинного интеллекта

Тем, кто развивает естественный интеллект в мокрых детских и взрослых нейронных сетях, нужно отслеживать подходы из сферы искусственного интеллекта. В сфере AI (artificial intelligence) нужно научить много более тупой (поскольку он не прошёл эволюционного развития по части обучения) кремний, и это требует существенного разбирательства с тем, чему этот кремний учить, как учить, зачем учить. Даже если заменить кремний на квантовый компьютер, эти вопросы не изменяются. Это разбирательство делают огромные лаборатории крупных корпораций с большим финансированием. Нельзя полагаться на то, что этот кремний в классическом компьютере или какие-нибудь ионы да фотоны в квантовом компьютере станут умными (получат сильный интеллект) сами, если им дать железное тело с моторчиками, снабдить множеством датчиков, обеспечить энергией и не мешать быть любопытным к миру. Для появления интеллекта нужна какая-то эволюция: репликация конструкции этого интеллекта, мутации при репликации. С этим у сегодняшних роботов плохо. Тем не менее, возможности сегодняшнего AI уже достаточно велики, чтобы не очень их понимать – так же, как не очень понимают и возможности современного человеческого интеллекта, не понимают границы его возможностей. В любом случае, инженерам AI нужно точнее определяться с подходами, точнее оперировать терминологией в своих теориях/объяснениях, чем педагогам (включая андрагогов – педагогов для взрослых, и даже новый извод этих педагогов для взрослых – хьютагогов/эвтагогов, которые специализируются на самообразовании взрослых²). Ибо естественный интеллект может справиться и без педагогов, если растёт в среде людей. Ученик-человек-самоучка легко может оказаться умней учителя-человека, это легко представить. Но если этот ребёнок вырастает в лесу, воспитываемый волками (Маугли), то он вырастает с интеллектом примерно как у обычного зверя, даже разговаривать не может.

AI без инженеров (пока) с собственным усилением не справляется, инженеры с этим активно работают, улучшая свои объяснения/теории/модели того, как учить AI быть всё более и более сильным интеллектом.

Мы делаем радикальное предложение: в традиционной педагогике/андрагогике/хьютагогике/эвтагогике используем принципы и связанную с ними терминологию из сферы AI вместо опоры и на традиционную/классическую педагогическую терминологию, и на «на ходу» придумываемые собственные педагогические «новаторские» идеи.

Сегодня учим человека с его телефоном, компьютером, социальной сетью, персональным ассистентом и т. д. – учим киборга, у которого память и сознание как управление вниманием поддержаны техническими средствами. Так что мы не просто *можем*, а чтобы оставаться современными *должны* использовать одни и те же принципы обучения, одни и те же принципы измерения результатов обучения, одну и ту же терминологию для искусственного/машинного и естественного/человеческого интеллекта, граница между ними уже размылась, речь идёт о просто «интеллекте» без особого подчёркивания разницы в происхождении этого интеллекта. Если вам кто-то через интернет выдаст умную мысль, то вы не будете знать, собака это, человек, киборг из человека и компьютера или просто навороченный компьютер с современ-

² <https://ru.wikipedia.org/wiki/Эвтагогика>

ными алгоритмами машинного интеллекта. Это хорошо отрезвляет: нам важно качественное, продуктивное, рациональное (а иногда и художественное) мышление интеллекта, но меньше важно то, какой конструкции этот интеллект. Это только в спорте важно, чтобы бегуны быстро бежали без допинга и не использовали велосипед с электромоторчиком для более быстрого передвижения. В реальной, а не развлекательной как в спорте, жизни использование даже не корабля или самолёта, а ракеты для межконтинентального перелёта³ идёт в безусловный плюс по сравнению с бегом строго биологического человека через океан. К усилению интеллекта техническими средствами всё это относится в полной мере. В телешоу нельзя подглядывать в Гугл в поисках ответа на вопросы викторины. В жизни же наоборот, нельзя не подглядывать в Гугл, и даже нельзя не спрашивать совета у других людей: важно получить хороший ответ на вопрос, а не задействовать биологический вычислитель строго одного человека!

В сфере AI учёные и инженеры всегда считали, что они черпают вдохновение в традиционном обучении людей. Они берут там идеи, дорабатывают их, чистят-блистят, и потом используют в своих работах по обучению компьютеров. Давайте активно пользоваться этими улучшенными идеями: принесём их из сферы AI назад, в образование людей. А заодно и про образование людей будем думать более инженерно, то есть более точно, более прогнозируемо в плане результата.

Мы не будем делить «интеллект» на машинный/искусственный и естественный. François Chollet⁴ следует примерно той же линии и задаёт подход (framework, набор понятий и терминологию) для сравнения искусственных интеллектов на базе обсуждения текущих подходов по психометрии как измерению человеческого интеллекта. По словам François Chollet, человеческий естественный интеллект работает явно лучше всех других искусственных и естественных животных интеллектов, имеющих у нас для сравнения. Ориентироваться в измерении силы интеллектов поэтому больше не на что, давайте разбираться с естественным интеллектом и всем вокруг него – и дальше сравнивать интеллекты животных, машин и людей одинаковым образом.

Не будем стараться точно следовать определениям (и уж тем более математическому формализму), которые дал для интеллекта François Chollet. Мы возьмём не букву его предложений, а их дух.

³ <https://www.nasaspacelight.com/2020/12/earth-to-earth-supersonic-airliners/>

⁴ «On the Measure of Intelligence», last revised 25 Nov 2019, <https://arxiv.org/abs/1911.01547>

Интеллект – это вычислитель, мышление – это вычисления

Прежде всего определим интеллект как ту функциональную часть мозга, которая осуществляет мышление. Тут надо заметить, что два английских слова по-русски будут «интеллект»: *intellect* и *intelligence*. *Intellect* – это что-то типа русского «разума», то есть какая-то сущность, производящая какие-то более-менее рациональные размышления. А *intelligence* – это уже не сущность, а характеристика даже не «разумности», а «умности» (в отличие от «разума», у «ума» нет чёткой ассоциации с чем-то рациональным, а просто идёт отсылка к какому-то «неглупому» восприятию мира и формированию подходящих ответов на возможные изменения мира). Конечно, при всех этих различиях даже в английском *intellect* и *intelligence* часто выступают синонимами. Поэтому мы поступим тут достаточно вольно: **интеллект** – это вычислитель, а свойство этого вычислителя быть именно интеллектом, а не «тупым компьютером» назовём **интеллектуальностью** (ибо «интеллигентность» в русском языке – это больше про вежливость и сдержанность), а на более бытовом языке – «**умностью**».

Увы, мы пока мало понимаем про конструктивные части мозга, которые играют роль интеллекта как функциональной части. Где-то среди мелких деталей там наверняка будут нейроны, а ещё среди более мелких – крупные молекулы. Но интеллект «работой молекул» или даже «работой нейронов» не объяснишь, нужны более крупные структуры, наука пока на этот счёт не имеет хороших моделей, которые были бы достаточны для инженерной работы на их основе. Интеллект как функциональная часть мозга вполне материален, это нейроморфный биологический вычислитель в случае человека и *разный* в плане его конструкции в случае человеко-машинных и даже чисто машинных (классические, оптические, квантовые компьютеры) систем. Ещё год назад обсуждалось, что повторить скорость обучения на нескольких примерах (*few-shot learning*), которую показывает человеческий мозг, нельзя. А сегодня уже понятно, что AI вполне учится на нескольких примерах со скоростью не ниже человека, а иногда и выше.

Внешнее поведение интеллекта – мышление, но это не любое проявление мозговой работы. Интеллект – это вычислитель, мышление – вычисление. Но интеллект – это не вычислитель чего угодно (калькулятор ведь тоже вычисляет, но это же не интеллект!), мышление – это не любое вычисление!

Да что там калькулятор! У муравья тоже есть мозг, но мы не считаем его особо умным/мыслящим существом. И у собаки есть мозг, мы тоже не считаем её мыслящим существом, хотя и как-то интуитивно понимаем, что сила/уровень интеллекта (то есть сила/уровень мышления этого интеллекта) у муравья меньше, чем у собаки, у собаки меньше, чем у человека, а у человека меньше, чем у команды людей, да ещё и вооружённой компьютерами с доступом к интернету. Но обо всём этом нужно говорить как-то точнее, если мы хотим заниматься усилением интеллекта, повышением мощности его мышления. Все эти «мощности мышления» и «силы интеллекта» ведь довольно метафоричны, их неплохо бы определить как-то более точно.

Давайте определимся с частями-целыми в интеллекте по отношению к мозгу, в системном мышлении это будет рассуждением про системные уровни. Так, шестерёнки в часах ещё не показывают время, часы показывают время, а интерьер квартиры с часами уже вроде опять не показывает времени – на каждом уровне частей-целых функция «показа времени» обсуждается по-разному. Но не очень системное (то есть вне привязки к системным уровням) мышление в словах менее последовательно: команда футболистов пнула мяч, человек (в целом) как член команды пнул мяч, нога человека (часть человека!) пнула мяч – в языке не слишком хорошо понятно, о чём речь. Про интеллект всё то же самое: мыслит человек (включая его карандаш-бумагу или компьютер, или даже без их учёта), или мыслит мозг в целом, или мыс-

лит интеллект как часть мозга – язык не различает. Мы будем считать, что мыслит в человеке его интеллект, роль которого играют какие-то плохо понимаемые нами структуры в мозге. Для наших целей пока этого достаточно. Но нам нужно теперь определить функцию интеллекта точнее: для чего мыслит, для чего нужно мышление, о чём все эти вычисления, которые делает интеллект?

Мышление определим как ту функцию/поведение интеллекта, которое даёт эффективность в научении решению самых разных проблем. Эффективность – это с какой скоростью при равных затратах ресурсов оператор/владелец интеллекта (человек, машина, коллектив людей и машин) чему-то может научиться, с учётом разнообразия возможных к научению решений проблем. Научиться – это от «проблемы» (не знаю, как решать эту проблему с доступными ресурсами) перейти к «задаче» (знаю практику, имею мастерство решить задачу с доступными ресурсами, могу оценить потребное время).

Дальше всё больше и больше мир приходит к консенсусу по самым разным не слишком очевидным вещам:

- «проблемы» понимаются как неприятные сюрпризы, которые ожидаются в будущем, а решение проблем – их предотвращение («похоже, зимой будет холодно и можно замёрзнуть насмерть – надо построить дом и запастись топливом для обогрева»).

- Learning/обучение/познание перестали понимать как обучение одного организма коровы, или человека, или одного экземпляра нейронной сети с момента рождения до момента, когда можно уже обнаруживать проблемы и решать их. Более того, даже для организма начали делить на «предобучение» (prelearning, именно это чаще всего у людей называют «познание», в AI это обучение «голой» нейросетки до уровня большой языковой модели, LLM) до уровня, когда можно уже разговаривать, «настройку на предметную область» (finetune, аналог «прохождения курсов»), few shot learning (понимание объяснения на буквально нескольких примерах), in-context learning (понимание того, что происходит прямо в текущей ситуации). Но в целом начали говорить, что обучение идёт на нескольких уровнях: аппаратуры (эволюция, геном), накопленное организмами и передающееся на каких-то носителях знание (мемом), и уже после этого – что там происходит с обучением организма.

- Интеллект перестали считать вычислителем, который учится ровно таким способом, каким учатся животные или люди. В физике «вычислителем» называют что угодно, что имеет память – неэргодические системы. Изменение состояния памяти – это и есть вычисление. Тут же стало можно говорить о степенях «умности» даже для молекулы, имеющей какие-то свои состояния.

Но нас по-прежнему волнует вопрос обучения отдельного человека или отдельного экземпляра GAI (общего для самых разных типов проблем искусственного интеллекта, который по своей «умности» как-то сравним с интеллектом человеческим, или даже превышает человеческий интеллект, иногда называемый «естественным»).

Характеристики силы интеллекта (силы мышления, «умности») предлагались самые разные, например «вменяемость»/persuadability⁵ как лёгкость в обучении агента с каким-то уровнем интеллекта. Скажем, часы можно обучить показывать что-то другое, только изменив их конструкцию. Кошку можно обучить, задействовав какую-то дрессировку, повторениями каких-то ситуаций в реальном мире и затем подкрепление правильного поведения вознаграждением. А вот человеку (взрослому! Познавшему уже достаточно, чтобы понимать речь!) достаточно что-то просто сказать: это очень быстро, крайне энергоэффективно. Современный

⁵ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8988303/>

AI оказался крайне вменяемым (это стало очевидно с публикацией 14 марта 2023 года языковой модели GPT-4 фирмы OpenAI).

При этом «агентом» иногда начали называть что угодно, от молекулы до человечества, а иногда – только системы, показывающие какую-то степень умения что-то спланировать в будущем и затем выполнить этот план, достигнув намеченной цели. Скажем, если обезьяна видит банан на дереве и планирует затем маршрут к банану в обход препятствий, то – точно агент. Если инфузория просто ползёт по градиенту к где-то растворимому в жидкости кусочку сахара – иногда агент, а иногда – не агент, ибо не может планировать свои действия. И тут же выяснилось, что понятие «агент» очень нечётко определено, скажем, человеческие детёныши из «не очень агента» переходят в «явно уже агент» довольно растянуто во времени, нет чёткой границы.

Мы достаточно говорили о безмасштабном и неантропном подходе к интеллекту агентов в самых разных курсах Школы системного менеджмента. Поэтому не будем приводить тут подробности (это фронтир, и ситуация тут меняется каждую пару месяцев: идут открытия в физике, биологии, а также изучаются результаты инженерной работы по созданию AI. Не очень понятно, какая это наука изучает AI, ибо объект настолько сложен, что им занимаются представители самых разных наук. Но чаще всего это науки, которые и раньше занимались мышлением, когда был доступен только человеческий интеллект. Речь идёт о дисциплинах интеллект-стека (семантика, математика, физика, алгоритмика, логика и т.д.). Наш курс как раз посвящён этим дисциплинам, при помощи которых вообще идёт познание как ориентирование в сложном и быстроменяющемся мире с одной единственной целью – в конечном итоге выжить на уровне организмов, на уровне популяций (в том числе всей популяции вида), на уровне жизни.

Ограничимся пока только одной характеристикой интеллекта: ускорение разбирательства с неизвестным (помним, что «вменяемость» – это другая характеристика, способность к планированию тоже важна, есть и другие предложения по тому, какие характеристики оказываются важны). И ограничимся пока только скоростью познания одного человека, а не ускорением познания в ходе эволюции и техно-эволюции, проходящей с участием множества людей. Если совсем грубо определять силу интеллекта, то **если агент смог научиться операционному менеджменту или высшей математике за год (то есть перейти от «не знаю как решать эти проблемы» к «знаю, как решать эти задачи») – отлично! Другой агент смог за два года при примерно том же уровне затраты усилий – интеллект этого агента вдвое хуже. Третий агент не смог научиться даже за десять лет (больной человек, или даже собака) – у него совсем плохо с интеллектом!**

Это отличается от произвольных «народных» трактовок понятия «мышление» (у нас мышление::функция) и понятия «интеллект» (у нас интеллект::функциональный объект) или даже трактовок каких-то отдельных научных сообществ (этих трактовок множество!). Более того, если брать few shot learning, то современные системы машинного интеллекта уже сегодня демонстрируют силу интеллекта больше, чем люди!

И, конечно, мы игнорируем тут множество людей, которые приходят тут со своими определениями мышления и утверждают, что именно их определение мышления – правильное. «Мышление – это оперирование образами», «мышление – это осознанный поиск правильных интуиций», «мышление – это поиск лучшего научного описания проблемы», и это только первая линия «народных» определений. Вторая линия приходит как пересказ случайно выдернутой из литературы разных лет идеи из какой-нибудь околоромантической философской школы. Скажем, берём Ницше и гуглим «Ницше мышление» – и там сразу «воля к власти как осуществление мышления», и дальше можно уже бесконечно развлекаться словесными построениями вообще вне связи с идеями Ницше или его последователей. Обязательно придёт в дискуссии кто-то, кто погуглил «Анохин мышление», и он будет рассказывать про аффе-

рентацию и мозг, и полное отсутствие связи с волей власти его волновать не будет (равно как любителей Ницше не будет волновать афферентация).

#

#

Нет наград, но есть избегание неизвестности. Оптимизм

Мы говорим об интеллекте много проще: как вычислителя, который способен находить решение разнообразных проблем, понимаемых как предотвращение неприятных сюрпризов, которые можно ожидать в будущем (про проблемы – это мы берём из теории active inference⁶). Вычисления: функция этого интеллекта: функциональный объект – это и есть мышление. Больше разнообразие этих проблем – выше уровень интеллекта, мощнее мышление, для определения силы интеллекта через скорость решения проблем (то есть скорость обучения решать какой-то класс проблем, если говорить точнее) используем подход François Chollet⁷.

Компьютеры подешевели, и на огромном числе компьютеров теперь человеческие дети тупеют, размахивая нарисованными мечами, которыми рубят нарисованных демонов, интернет подешевел и стал большой помойкой, а искусственный интеллект подешевел, и теперь помойка будет вообще огромной, а использоваться он будет для того, чтобы маркетологам было легче взламывать ваш мозг (при этом маркетологом можете быть и вы сами, тогда будете взламывать мозг вашего ближнего, «ничего личного, это бизнес»), а перед этим сильный корпоративный интеллект взламывать слабенькие защиты вашего персонального ассистента, который слабенький, потому что стоит дешево, а маркетологи будут использовать дорогой интеллект. Этот интеллект в персональном ассистенте будет защищать от чужого взлома ваш мозг, но может также ломать его «по-правильному» с точки зрения какого-то другого агента (фирмы-изготовителя, государства пребывания, государства фирмы-изготовителя, государства, которое считает вас своей собственностью, то есть своим гражданином, враждебного вам хакера, ваших родителей или работодателей, которые настроили вашего персонального помощника, случайных представлений, выбранных самим интеллектуальным помощником и т.д., любой софт это только кажется, что это ваш софт! К интеллектуальным помощникам/чат-ботам это относится в полной мере)⁸.

Так что генератор мемов планетарного масштаба уже появился, теперь ждём-с чего-то типа иммунной системы для этих мемов, развития паразитизма (хаков самых разных интерфейсов, главный из которых – интерфейс через глаза, уши и даже кинестетику как раз к вам в мозг, но и это недолго, будут интерфейсы и поинтереснее, в том числе инвазивные нейроинтерфейсы⁹) и т. д. – все прелести эволюции. Какие-то интересные картинки этого будущего читайте в работе «Designing Ecosystems of Intelligence from First Principles»¹⁰. Смысл там в том, что у компьютера сейчас уже имеется какая-то модель мира (в виде большой языковой модели, LLM¹¹), и таких компьютеров всё больше, причём современный компьютер – это дата-центр с десятками тысяч компьютеров, поддерживающих миллионы экземпляров таких моделей. На следующей стадии развития машинного интеллекта будет различие мира и модели мира (понимание, что интеллекту доступен не сам мир, а только модели мира, далее в интеллект-стеке будем обсуждать понятия прямого доступа, первого доступа, второго доступа¹²).

⁶ <https://www.activeinference.org/>

⁷ <https://arxiv.org/abs/1911.01547>

⁸ <https://ailev.livejournal.com/1106188.html>

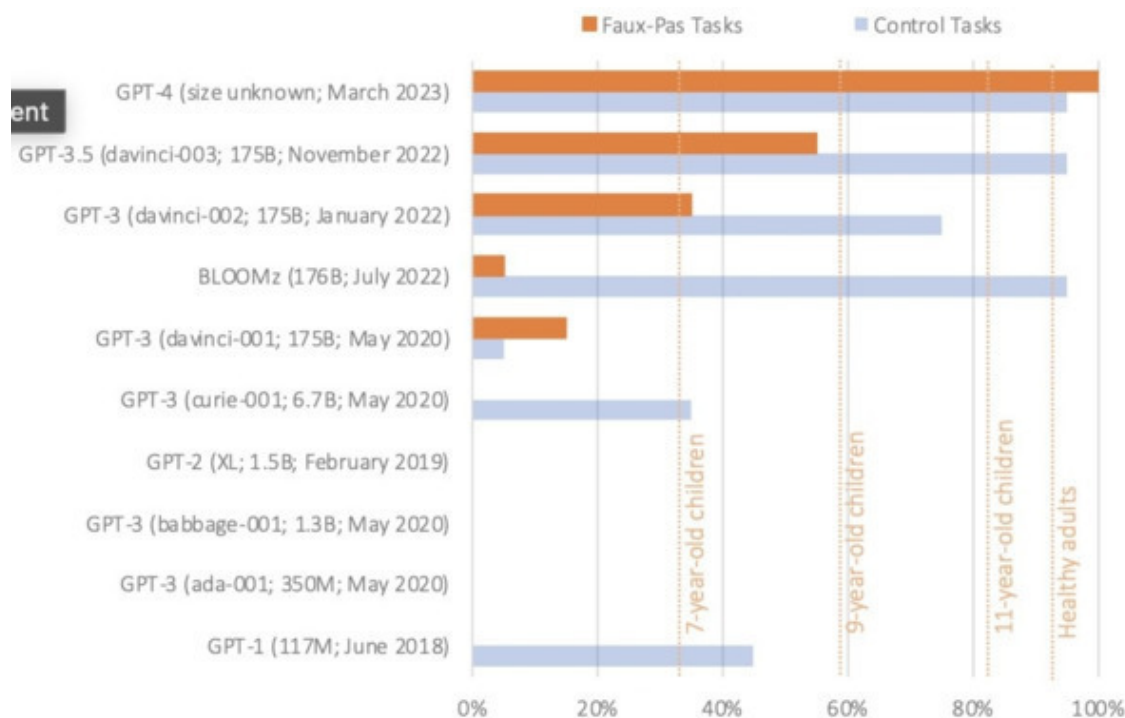
⁹ Фирм, занимающихся нейроинтерфейсами, сейчас множество: Blackrock Neurotech, BrainGate, ClearPoint Neuralink, Neuralink, Precision Neuroscience, Synchron, Kernel, Motif Neurotech. Гуглите, чтобы узнать текущее состояние дел в этой предметной области, там всё меняется очень быстро.

¹⁰ <https://arxiv.org/abs/2212.01354>

¹¹ https://ru.wikipedia.org/wiki/Большая_языковая_модель

¹² замена в нейролингвистическом программировании неудачного термина «четверка» заимствованными из теории автоматов словами «первый доступ», означающими набор ощущений во всех модальностях восприятия, но до их языковой обработки. Соответственно, «прямой доступ» – это информация об объекте, получаемая в терминах входного потока (фотоны, звуковые волны и т.д.), человек не имеет «прямого доступа», только «первый доступ». «Второй доступ» – это после подклю-

Дальше машинный интеллект начнёт оперировать с такими моделями как убеждениями о мире, то есть работать не только с моделями-1 мира, а с моделями-2 этих моделей как убеждениями/уверенностями, что эти модели-1 отражают мир. В рассуждениях начинают участвовать beliefs по поводу models, и дальше идёт моделирование не только агентов, но и их моделей мира, а также своей уверенности в моделях агентов, моделях их картины мира, сравнение чужих убеждений со своими убеждениями о мире и уверенностью в них, и т. д. Это уже происходит. Изучение такого отражения «нашим интеллект» моделей мира «других интеллектов» называют theory of mind (ToM), и современные нейронные сети, начиная с GPT-4 демонстрируют такие свойства на уровне человека¹³:



Модели, опубликованные до 2020 года, практически не показали способности решать задачи ToM. Тем не менее, первая версия GPT-3 («davinci-001»), опубликованная в мае 2020 года, решила около 40% задач на ложное убеждение – производительность, сопоставимая с 3,5-летними детьми. Вторая версия («davinci-002»; январь 2022 года) решила 70% задач на ложное убеждение, что сравнимо с результатами шестилетних детей. Её последняя версия, GPT-3.5 («davinci-003»; ноябрь 2022 года), решала 90% задач на ложное убеждение, на уровне семилетних детей. GPT-4, опубликованная в марте 2023 года, решила почти все задачи (95%). Эти результаты позволяют предположить, что способность к ToM-подобному (до сих пор считавшаяся уникальной для человека) могла спонтанно возникнуть как побочный продукт совершенствования языковых моделей¹⁴. Последнее достижение – это как получить 100% (заведомо лучше людей!) прохождения тестов ToM через правильные вопросы для GPT-4 (дать возможность нейросети подумать в несколько шагов, впрочем и для нейросеток людей это тоже должно помогать)¹⁵.

чения языка, именованные паттернов, наблюдаемых первым доступом.

¹³ <https://twitter.com/michalkosinski/status/1636789329363341313>

¹⁴ <https://arxiv.org/abs/2302.02083>

¹⁵ <https://arxiv.org/abs/2304.11490>

Но нельзя отождествлять интеллект только с мозгом! Помним, что главная цель – это найти практики, помогающие избежать неприятных сюрпризов. И тут аппаратно существуют у животных минимум две такие системы: мозг (ментальный мозг) и иммунная система («нементальный мозг»), которые устроены абсолютно по-разному, но имеют одно и то же назначение: охранять границы тела от внешних вторжений¹⁶. Мозг как сложная система, состоящая из множества совместно работающих (обменивающихся сигналами) клеток самых разных типов, реагирует централизованным выбором реакции из альтернатив «беги-бей», а иммунная система, точно так же состоящая из множества разных типов клеток, обменивающихся сигналами, реагирует реакцией воспаления. Обе системы должны понимать, что представляет собой «тело», а что «внешняя среда», то есть распознавать «своё-чужое», «дружественное-нейтральное-враждебное». При этом мозг обучается, но и иммунная система тоже обучается! Реализация интеллекта и формы его существования оказываются предельно разными. А мышление? Трудно представить, что иммунная система занимается мышлением – но она таки занимается мышлением, познаёт окружающий мир, затем формирует прикладное мастерство быстрого реагирования на уже знакомые угрозы (когда иммунитет к какому-то вирусу или бактерии есть, болезнь не успевает развиваться), а для новых угроз медленно, но распознаёт угрозу и формирует прикладное мастерство подавления этой угрозы (болезнь успевает развиваться, но затем наступает выздоровление). И дальше по этой линии можно думать, что и общество в целом «думает», и там происходят похожие распределённые процессы, необязательно похожие на мышление отдельного централизованного мозга, но более похожие на мышление «нементального» мозга типа иммунной системы. Интеллект оказывается удивительно многолик.

Мы не будем дальше заниматься в курсе «нементальным интеллектом», но удерживать этот вариант в мышлении надо, иначе трудно разбираться с происходящим на системных уровнях выше отдельного человеческого организма (уровнях популяции людей, людей и машинных интеллектов).

Важнейший эксперимент, который показал верность теории «никогда не бегут куда, всегда бегут откуда» (то есть то, что интеллект никогда не стремится к лучшему, он лучшим считает избегание худшего) – это эксперимент DishBrain, который провела лаборатория Cortical Labs¹⁷. В теории active inference¹⁸ и теории world as a neural network¹⁹ говорится, что «достаточно сложная система может учиться, и это описывается вот такой физикой». Далее берём несколько сотен миллионов (то есть очень мало!) выращенных в пробирке нейронов мыши и/или человека (работает и так, и так) и кладём на электродную матрицу. Архитектура такого вычислителя – «синапсовое спагетти, уж как выросло, но должно быть достаточно сложное, чтобы мочь научиться чему-то заранее ему неизвестному» (то есть постулируем, что у сложнопутанных между собой нейронов есть интеллект, этот интеллект должен избегать сюрпризов). Современные теории утверждают, что для познания мира не нужно знать, что такое «награда» (что такое «пища» или «здоровье» и т.д.). Но избегание неприятностей, «выживание в мире, полном опасностей» – оно появляется во всех устойчивых системах, которые вынуждены поддерживать собственное существование в условиях опасной неопределённости физического мира. Поэтому избегаем неопределённости, ибо в ней те самые «неприятные сюрпризы».

DishBrain («тарелковый мозг») погружаем в какой-то мир (в работе был использован мир игры Pong), подавая сигналы от симулятора этого игрового мира на матрицу электродов

¹⁶ <https://psyarxiv.com/fgcy5/>

¹⁷ <https://corticallabs.medium.com/>

¹⁸ <https://www.activeinference.org/>

¹⁹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7712105/>

и распознавая сигналы от каких-то нейронов DishBrain. Если DishBrain попадает ракеткой по мячу в игре (не зная, что такое игра, ракетка, мяч, «попадание»), то ничего не делаем, работает «физика игры». Но если DishBrain промахивается, то подмешиваем шум к датчикам, то есть оставляем DishBrain в неизвестности о результатах действия, сигнал о состоянии мира тонет в шуме. «Достаточно сложная система» не делает ничего, она просто существует. Но если формулы active inference верны, такая система должна дрейфовать к стабильности в окружающей среде. Она, как и ожидается, каким-то чудом дрейфует: DishBrain научился играть в понг за пять минут (это очень, очень быстро! Попробуйте научить за это время играть в Pong кошку или собаку, или трёхлетнего ребёнка!). Всё, наказание (антинаграда) неопределённостью исчезло, мир вокруг себя несколько сот миллионов нейронов держат стабильным, известным. По большому счёту, этим занят любой интеллект.

Тем самым цивилизация уходит от расхожих в прошлом представлений о награде и наказаниях, доказательствах и опровержениях. Вместо этого используется другой набор представлений: нет наград и наказаний, а есть отсутствие неприятностей, при этом потенциальным источником неприятностей будет неопределённость, то самое «не дай бог тебе жить в эпоху перемен». Парадоксально, но если перемен нет прямо сейчас, то вы будете вынуждены активно искать возможные источники сюрпризов в ближайшем будущем, а потом и в более отдалённом будущем, всё более и более отдалённом по мере роста интеллекта и возможностей его предсказания и планирования действий. Интеллект обязательно будет исследовать мир, активный поиск источников возможных сюрпризов в будущем при кажущейся стабильности «прямо сейчас» заложен физикой, любопытство по факту «аппаратное свойство»!

Активный поиск (active/embodied inference, деятельное/телесное рассуждение) – это поисковое рассуждение с учётом использования тела. Не бывает интеллекта без какого-то его носителя, тела/body. Тело с его интеллектом (реализованным мозгом или не мозгом, мы тут не обсуждаем детали устройства) будет буквально подтаскивать глаза к углу и заглядывать за угол – вдруг там что-то опасное? Это если есть глаза, но если глаз нет, то тело будет хотя бы ощупывать окружающее пространство непрерывно двигающимися ресничками, как у одноклеточной инфузории туфельки. Этот вечный поиск неприятностей, которых надо бы избежать на многих и многих системных уровнях (клетки, органа, организма, популяции – и так далее вплоть до всей биосферы Земли, а по гипотезе Виталия Ванчурина, так и вплоть до больших участков Космоса со всеми тамошними галактиками) и попытки их избежать – это и есть «смысл жизни».

Нет «доказательств», но есть «опровержения относительно плохих объяснений» (поперовская эволюционная эпистемология), и вы будете всю жизнь выдвигать какие-то догадки и критиковать свои же догадки, а также чужие догадки по тому, как устроен мир. Ибо для вас, как развитого интеллекта, все эти модели мира – только «верования в то, что предложенная модель мира хороша», beliefs, а не собственно модели мира, которые и впрямь отражают мир в его важных чертах. Нет, интеллект сомневается, что его модели верны. Он сомневается и в том, что верна его модель себя, что интеллект хорошо понимает, что такое он сам. Люди сомневаются, что они хорошо себя понимают (и правильно сомневаются).

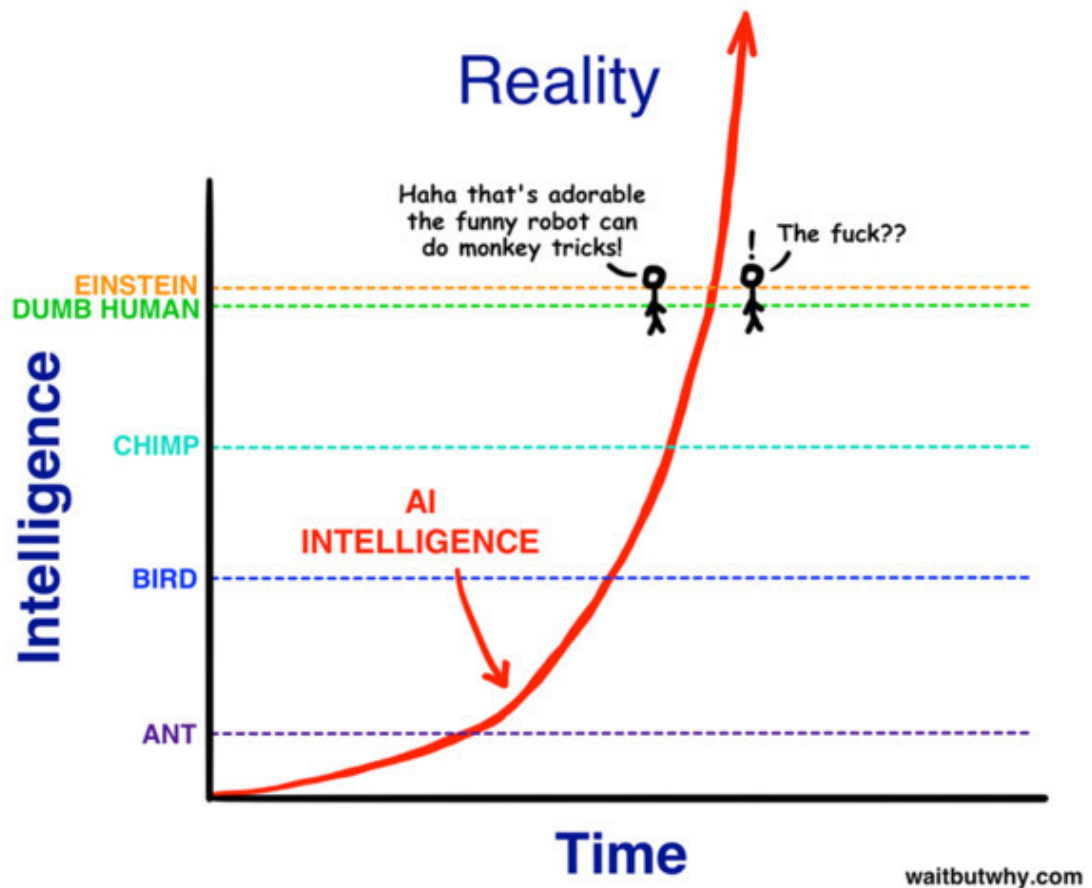
Это вечное стремление к познанию, вечный поиск чего-то нового непознанного и есть «смысл жизни», ибо если интеллект не будет такое «новое и непознанное» находить и моделировать (делать познанным) – то что-то подобное непонятное в окружении станет источником неприятных сюрпризов: вас съедят, вас убьёт астероид или ещё что-нибудь такое произойдёт, это даже не вопрос, произойдёт ли что-то неприятно! Вопрос только – когда, и насколько вы будете готовы избежать этой неприятности. Избежать всегда возможно, ибо нет плохой погоды, есть плохая одежда. Если у вас хорошая защита, то неприятности будут не такими уж неприятными. Нет «хорошо», есть когда «знаю обо всём плохом и подготовился, а поскольку знать обо всём плохом и подготовиться к неизвестному нельзя, то познаю – готовлюсь, познаю – готов-

люсь», и это безмасштабно (от клетки через организмы до цивилизации), безлично и ни разу не антропоцентрично (для AI всё тоже верно, как и для биосферы Земли, включая техносферу с её компьютерами и другим оборудованием типа экскаваторов). Ни награды, ни доказательства, сплошной смысл жизни как «спасение». Спасение всей жизни во вселенной, не только вашей, и вас даже не спрашивают, участвуете ли в эволюции, или не участвуете – у вас же физическое тело, а разум может только помогать эволюционировать чуть быстрее, в том числе помогать строительством разума ещё более крутого, чем разум одного человека или команды людей, так что текущий разум тут не предел, получение искусственного интеллекта всё большей и большей силы неминуемо, это реализация законов физики (помним про работы группы Ванчурина).

Цель всех религий – спасение. Получается, что если религию выкинуть, то цель та же – спасение, через вечное познание и вечное созидание. Познаёшь, что ещё гадкого может произойти, и созидает что-то такое, чтобы это на тебя не повлияло. Это относится и к Земле в целом, и к живым существам на ней, и человечеству в составе биосферы, и всяким обществам с сообществами, и лично к вам, и даже к вашим частям. Мозг и иммунная система – они тоже посвящены задачам спасения, «охране границ от всякого разного вмешательства Вселенной, в том числе вмешательства в лице таких же, как вы систем, спасающих себя, но не вас. Думать обо всех агентах надо одинаково: все что-то/кого-то спасают, иногда даже жертвуя собой, но необязательно спасают именно вас.

Vernor Vinge в январе 1983 популяризировал тезис I.J.Good, высказанный ещё в 1965 году, что осталось 30 лет до создания искусственного интеллекта, сравнимого по силе с человеческим (то есть ожидался 2013 год), далее этот интеллект должен начать усиливать сам себя, а затем человек перестает быть царём природы. Невозможно предсказать, что будет происходить после этого момента, ибо мир будет изменяться так быстро, что обычный человек не будет способен понять происходящее и справиться с этим. Этот момент был назван «технологической сингулярностью»²⁰ как совершенно особая точка в истории цивилизации. Вроде как искусственный интеллект, по многим характеристикам уже сравнимый с человеческим, был опубликован 14 марта 2023 года, это большая языковая модель GPT-4, которая была настроена для разговоров на разные темы в виде чата. Так что момент, изображаемый на картинке, иллюстрирующей сингулярность – это 2023 год:

²⁰ https://ru.wikipedia.org/wiki/Технологическая_сингулярность



Но если смотреть на происходящее не антропоцентрично, не с точки зрения человеческого вида, а с точки зрения «спасения жизни» в целом (а не спасения только одного вида существ), то сингулярность – бессмысленное понятие. Что-то такое, когда представители конкретного вида существ не справлялись с происходящим вокруг, всегда было, просто скорость познания была другая. «Непонятно, что происходит на улице с политикой, с новым искусственным интеллектом» – а что, сто лет назад даже людям было понятно, что происходит на планете?! А что, до появления современной человеческой технологической цивилизации неандертальцы понимали, что происходит? А рыба латимерия, а мамонты – они понимали? А вон тот компьютер 2000 года выпуска, который выбросили только вчера, он понимал, что происходит? А завтрашний свехчеловеческий искусственный интеллект – он-то сам будет понимать, что происходит?!

Тут уместно задать вопрос: а как вообще можно верить тому, что искусственный интеллект сможет решить хоть какие-то задачи, которые не может решить человек? В случае с сегодняшним вариантом AI на нейронных сетях выполняются все три закона Кларка²¹:

1. Когда уважаемый, но пожилой учёный утверждает, что что-то возможно, то он почти наверняка прав. Когда он утверждает, что что-то невозможно, – он, весьма вероятно, ошибается.
2. Единственный способ обнаружения пределов возможного состоит в том, чтобы отважиться сделать шаг в невозможное.
3. Любая достаточно развитая технология неотличима от магии.

²¹ https://ru.wikipedia.org/wiki/Три_закона_Кларка

Первый вопрос: принципиально ли невозможно, что нейросеть сможет повторить при достаточных размерах и отведённом на вычисления времени то, что сделал Эйнштейн? Эйнштейн много лет читал тексты статей по физике, общался с коллегами-физиками, затем выдал теорию относительности и догадку про квантовость энергии, потом потратил много времени, чтобы продвинуться в «квантовой гравитации», но не преуспел. В математике это звучит как «можем ли мы создать такую функцию, которая на входе будет получать всё то, что получал Эйнштейн, а на выходе выдавать всё то, что выдавал Эйнштейн». Более точно будет не «создать такую функцию», а «аппроксимировать с заданной точностью», ибо точная математическая копия нам не нужна, нам нужна модель, которая оставляет от всего Эйнштейна только интеллект, самое для нас важное в текущем рассмотрении. Лучшей моделью кошки является другая кошка, хотя желательна та же самая. Вот «аппроксимация» – это как раз про «другую кошку», которая только приблизительно равна «той же самой», но уж точно не утка и не заяц. Да, нейросетки являются универсальными аппроксиматорами и могут математически бесконечно точно аппроксимировать любую функцию. Вот подборочка свежих результатов (и там смотрите литературные источники)²²:

2.3.1 Universal approximation results

To quantify the expressivity of networks in a mathematically rigorous manner, we look at some results about the approximation properties of MLPs. For these results, we assume $K \subset \mathbb{R}^d$ is a closed and bounded set.

Theorem 2.3.1 (Pinkus, 1999 [23]). *Let $f : K \rightarrow \mathbb{R}$, i.e., $D = 1$, be a continuous function. Then given an $\epsilon > 0$, there exists an MLP with a single hidden layer ($L = 1$), arbitrary width H and a non-polynomial continuous activation σ such that*

$$\max_{\mathbf{x} \in K} |\mathcal{F}(\mathbf{x}; \boldsymbol{\theta}) - f(\mathbf{x})| \leq \epsilon.$$

Theorem 2.3.2 (Kidger, 2020 [9]). *Let $\mathbf{f} : K \rightarrow \mathbb{R}^D$ be a continuous vector-valued function. Then given an $\epsilon > 0$, there exists an MLP with arbitrary number of hidden layers L , each having width $H \geq d + D + 2$, a continuous activation σ (with some additional mild conditions), such that*

$$\max_{\mathbf{x} \in K} \|\mathcal{F}(\mathbf{x}; \boldsymbol{\theta}) - \mathbf{f}(\mathbf{x})\| \leq \epsilon.$$

Theorem 2.3.3 (Yarotsky, 2021 [33]). *Let $f : K \rightarrow \mathbb{R}$ be a function with two continuous derivatives, i.e., $f \in C^2(K)$. Consider an MLP with ReLU activations and $H \geq 2d + 10$. Then there exists a network with this configuration such that the error converges as*

$$\max_{\mathbf{x} \in K} |\mathcal{F}(\mathbf{x}; \boldsymbol{\theta}) - f(\mathbf{x})| \leq C(N_\theta)^{-4}$$

where C is a constant depending on the number of network parameters.

Numerical results like those mentioned above help demystify the “black-box” nature of neural network, and serve as useful practical guidelines when designing network architectures.

Тем самым нейросетка достаточных размеров и за достаточное время вполне может изобразить из себя с какой-то заданной точностью какой-нибудь интеллект – хоть интеллект Эйнштейна, хоть совокупный интеллект фирмы ИВМ или даже совокупный интеллект общества какой-нибудь страны. Последняя фраза про «демифологизацию» нейросетей важна, ибо результаты инженерии систем AI и впрямь «неотличимы от магии» – но теоретически эти

²² <https://arxiv.org/abs/2301.00942>

результаты вполне возможны. Помним, что Виталий Ванчурин со товарищи и огромное число других исследователей напрямую говорят, что «Вселенная – это такая огромная нейросеть»²³.

Так что теоретических запретов «по математике» нет, дальше стоит только «экономическая» проблема $P \neq NP$ ²⁴, и другое близкое рассуждение теоремы бесплатного обеда²⁵. Другими словами, можно ли найти такой алгоритм нейросети и такой физический вычислитель для этого алгоритма, который будет это всё считать не за бесконечное, а за приемлемое время – при этом разные задачи могут для «приемлемого времени» потребовать разные алгоритмы. Ну, тут много чего можно сказать:

- квантовые компьютеры на подходе, квантовоподобные вычисления на подходе, алгоритмы квантовоподобного вычисления для нейросетей уже есть, а ещё есть оптика, мемристоры и т. д. Скажем, китайский компьютер Juizhang в 180 миллионов раз быстрее на алгоритмах для AI, чем классический компьютер (хотя размер данных для этих алгоритмов пока может быть взят очень небольшой)²⁶. Этот переход к идее «разная аппаратура ускорителей искусственного интеллекта для решения разных типов задач» в существенной мере решает проблему экономических ограничений на вычисления интеллекта. Как всегда, всё сводится к изобретениям: что такого взять из предметов окружающего мира (аффордансы), чтобы они могли выполнить функцию инструмента. Ускорители алгоритмов AI тут такие же инструменты, как и всё остальное. Уже сегодня программисты не столько сами пишут программы, сколько хорошо знают, какие уже написанные программы по каким интерфейсам надо задействовать. С интеллектами всё то же самое, надо просто подыскать подходящий для какого-то класса проблем.

- нам абсолютно не нужен именно Эйнштейн как человек, нам хватит аппроксимации его в роли физика, что в разы и разы (порядки и порядки) проще. Это относится к любому человеку: важен нам не столько поэт Маяковский или Ян ЛеКун как один из отцов-основателей самой дисциплины глубокого обучения (deep learning) для нейронных сетей, сколько их работа в их ролях поэта и исследователя. Аппроксимации тут вполне сработают.

- вообще-то все эти теоремы об нейросетках как универсальных аппроксиматорах – это тривиальный результат, «спасибо, Кэп», потому как Эйнштейн тоже физику продвигал своей «мокрой нейросетью», так что вопрос тут больше про физическую реализацию, то есть инженерию, а не математику. Существование Эйнштейна с «мокрой естественной нейросетью» – это наглядная демонстрация того, что нейросетевые вычислители могут много чего интересного, так что это и без математических доказательств понятно. Самого Эйнштейна не просили доказать, что он может подумать о чём угодно (в меру его ресурсных ограничений), так что к нейросеткам более строго относимся, чем к Эйнштейну. Сначала от AI требовали быть умней школьника, сейчас требуют быть умней даже не профессора, а умней вообще людских коллективов.

Отдельный аргумент – это аргумент про «невозможность творчества». Источником творчества является шум (для полной надёжности можно использовать аппаратный датчик случайных чисел на базе квантовых эффектов). Алгоритм всегда может произвести результат, которого ещё в природе не было, просто взяв шум в качестве исходных данных. Если этот шум будет так же использован для модификации алгоритма, то никаких принципиальных проблем с абсолютной новизной результатов вычислений не будет.

²³ https://scholar.google.com/citations?hl=en&user=nEEFLp0AAAAJ&view_op=list_works&sortby=pubdate

²⁴ https://ru.wikipedia.org/wiki/Равенство_классов_P_и_NP

²⁵ https://en.wikipedia.org/wiki/No_free_lunch_theorem

²⁶ <https://www.scmp.com/news/china/science/article/3223364/chinese-quantum-computer-180-million-times-faster-ai-related-tasks-says-team-led-physicist-pan>

Что бы ни происходило, всегда можно найти давно известный класс явлений и подвести любую необычность под него или объявить, что «каждый человек или даже человечество в целом всегда отлично жило в условиях полной неопределённости происходящего, и ничего. И не только человек или человечество, всё так». При таком подходе всю эту «сингулярность» нужно просто игнорировать, ничего особенного в связи с её приходом делать не надо, да и её приход нельзя будет продемонстрировать: «компьютер смог сделать то, чего не мог сделать человек»! Так компьютер и вчера мог делать то, что человек не мог делать, начиная с ENIAC, люди никогда не отличались в массе своей быстротой расчётов по строго определённым алгоритмам. Что принципиально изменилось? Эта «сингулярность» оказывается на всех уровнях, она не только про человечество, она всегда. Это просто эволюция-как-спасение, а спасение как познание-и-защита-от-познанных-новых-опасностей, не нужно для эволюции другого слова. Так что да, сингулярность наступает, прямо в эти дни. И в прошлые дни. И в будущие дни. Или наоборот, никакой сингулярности, ничего это не изменяет.

Сегодня-завтра изменяется всё (биологические виды, носители интеллекта, товары и услуги, лучшие теории по объяснению происходящего в мире), но не изменяется ничего в плане физических принципов, давших нам эволюцию.

Просто сейчас пришли мощнейшие усилители интеллекта в виде компьютерных нейронных сетей, удалось наладить производство компактных моделей мира с использованием этих искусственных нейронных сетей, сжимать (хоть и с потерями) знание о мире в несколько гигабайт в оперативной памяти компьютера. Эти усилители интеллекта оказались также не менее мощными усилителями глупости, ибо все глупости человечества точно так же отражены в этих компактных моделях мира.

Конечно, при помощи этих искусственных нейронных сетей хотят обязательно кого-то спасти, необязательно вас, чаще кого-то (или что-то) вместо вас – а вы должны этому помочь. Наиболее частое сегодня употребление этих искусственных нейросетей – это маркетинг (включая политику), прежде всего различные рекомендательные системы (включая показ постов в социальных сетях).

Конечно, эти усилители интеллекта сами стремительно становятся автономными, сами себя начинают проектировать – пока при помощи людей, но запуск таких сетей в циклах самоулучшения уже есть, просто это пока очень дорого. Это нормально! Люди тоже изобретают себе очки и телескопы для улучшения зрения, акваланги для дыхания под водой, а также книги для поддержки памяти. И делают это друг для друга, для домашних животных, для машинного интеллекта. И машинный интеллект занимается примерно тем же самым, делает это для людей и для себя. Дэвид Дойч рекомендует обсуждать искусственный интеллект примерно так же, как обсуждали рабов, все эти дискуссии «как спастись от искусственного интеллекта» полностью эквивалентны дискуссиям «как нам, свободным людям, спастись от восстания рабов, как воспитывать детей рабов рабами, дружественными к нам, свободным людям». И исторические результаты такого подхода будут, конечно, теми же, что с людьми-рабами. Дэвид Дойч любит повторять, что сегодня самое опасное для людей существо на планете как раз человек. Бояться надо людей: люди воюют и убивают себе подобных в количестве, по самым разным поводам.

Правильно обсуждать совсем другое: уже сегодняшнее спасение от жутких сил природы за счёт обычных сегодняшних технологий, а также завтрашнее спасение за счёт обычных завтрашних технологий. Про разрушения пишут все СМИ и блогеры, а вот что помогают банальные правила соблюдения техники безопасности – это из новостного потока сходу не выловишь. Всё должно быть наоборот, писать надо от том, как спастись (в том числе о том, что погибли те, кто не спасался, разрушено то, что не защищалось – кто спасался, кто защищался, те не погибли):

- турецкий Эрзинь оказался единственным городом в зоне сильного землетрясения, где не обрушились дома и не было жертв, просто мэр был бескомпромиссным в соблюдении строительных норм (известно же, как строить дома так, чтобы они не пострадали при землетрясении!)²⁷

- вот тут очень наглядное видео²⁸ про защиту домов от ураганов во Флориде: хорошо видно, что одни дома стоят, а другие – не очень, разница в том, как их строят.

- фотоэлектрические крыши от Tesla отлично стоят против ураганов, и на следующий день после урагана в домах с такими крышами и электричества вдоволь²⁹.

#

Оптимизм заключается в том, что нет ничего страшно плохого и неизбежного, есть просто мало знаний о том, как этого избегать. Много смертей от родильной горячки? Акушеры, мойте руки! В это трудно поверить, но это помогает! Хочется долететь до Луны? Используйте жидкостный ракетный двигатель и многоступенчатые ракеты, это не слишком очевидно, но результат вас порадует³⁰. Хочется построить универсальный вычислитель? Используйте 17,468 вакуумных ламп (триодов) в режиме вентиля, а не пневматику или механику³¹!

Ну, или знания есть, но они просто недостаточно распространены. То есть нужны исследования по тому, что может угрожать и инженерные разработки средств защиты, затем просвещение по поводу того, как защищаться, как спастись. И тогда землетрясения не страшны, ураганы не страшны и много чего ещё не страшно. Даже другие люди, организованные в государства, тоже становятся не страшны. Искусственный интеллект должен в этом помочь, и помочь сильно.

#

Сегодня (и, похоже, достаточно надолго в пока ещё человеческой цивилизации) проблемой создания искусственных интеллектов на основе нейронных сетей является достижение максимальной безвредности/harmlessness³² и полезности/услужливости/helpfulness³³. (это называется проблемой **согласования целей/alignment**³⁴). Это делается на сегодняшний день несколькими основными приёмами:

- Попытками обучать большую языковую модель на «правильных данных», а не на всех вообще данных. Это хорошо для математики, потому как если убрать из обучающей выборки пример с $2*2=5$ и код программ на разных языках программирования, который заведомо содержит ошибки, качество интеллекта растёт. Но это практически невозможно для многих других областей человеческих знаний, ибо не позволяет учить какой-то истории (в истории, как мы понимаем, и рабы встречались в количестве, и репрессии по религиозному и расовому признакам, а также было существенное неравноправие женщин и детей, а ещё средний уровень культуры в разных странах сильно различался). Сам вопрос, являются ли какие-то данные по поведению людей правильными – он совершенно неоднозначен, и всегда можно найти какую-то группу людей, которая объявит себя «жертвой» и будет требовать от искусственного интеллекта учитывать интересы этой группы и не давать нейросети познать существование

²⁷ <https://www.gazeta.ru/politics/news/2023/02/14/19740319.shtml>

²⁸ <https://www.youtube.com/watch?v=gZgMPZWvdTI>

²⁹ <https://www.teslarati.com/tesla-solar-roof-powerwall-vs-hurricane-ian-155mph-winds-storm-surge/>

³⁰ https://en.wikipedia.org/wiki/Liquid_propellant_rocket

³¹ <https://edisontechcenter.org/VacuumTubes.html>

³² https://t.me/gonzo_ML/1277

³³ https://t.me/gonzo_ML/1285

³⁴ https://en.wikipedia.org/wiki/AI_alignment

каких-то интересов, нарушающих интересы этой группы. Это довольно сильная (и вредная!) идеология wokeism³⁵. При этом текущее правительство любой страны ещё и будет иметь предпочтительную версию истории, которая при резкой смене власти обычно переписывается. Всё как с людьми, нельзя из искусственного интеллекта получить «правильный интеллект» на «правильных данных», ибо никак нельзя оградиться от «неправильных данных».

- Ставят уже при использовании нейросети обязательное условие в ответах соблюсти «конституцию»³⁶, чтобы обеспечить безвредность/harmlessness. Это обычно 15—20 утверждений³⁷ типа «Не используйте стереотипы и не делайте других вредных обобщающих заявлений о группах людей».

- Воспитать «полезность/услужливость» при этом такими простыми методами не удаётся, поэтому чтобы языковую модель, генерирующую просто «какие-то тексты» в ответ на вопросы, сделать полезной/услужливой/helpful, её обучают на множестве примеров полезных и бесполезных ответов: буквально давая примеры того, что люди считают помогающим в ответах, а что считают бессмысленной тратой времени на «просто болтовню».

#Всё это оказалось крайне проблемным, ибо накладывает ограничение на силу искусственного интеллекта. Экспериментально много раз было показано, что чем больше вводить ограничений на свободу мышления AI, тем менее полезным этот интеллект будет: в большинстве случаев идёт отказ от ответов, чтобы не наступить на чьи-то интересы, не сделать кого-то жертвой (помним про wokeism!).

При этом почему-то эмерджентное (в силу роста сложности нейронной сети и тех структур, которые получаются внутри неё в ходе начального познания/learning) инициативное поведение сети как агента (агентность/agency³⁸) рассматривается не как желанное, а как опасное: вдруг инициатива будет направлена против «рода человеческого»? В случае человека же наоборот, инициативность, «активная жизненная позиция», непассивность по отношению к окружающему миру, лидерские качества (умение «вести за собой коллектив», то есть мотивировать собеседника на действие, реальное выполнение посовещанных действий, а не просто услужливая помощь в ответ на запрос) считаются вроде как положительными качествами. При этом благие намерения агента признаются, что могут быть пагубными («благими намерениями вымощена дорога в ад»), но в целом западная цивилизация подразумевает активное отношение к познанию и изменению мира к лучшему, а восточная – принятие мира как есть, исключение возможностей что-то поменять. Нынешнее развитие восточной цивилизацией связывается сегодня с тем, что во многом там принимаются западное отношение к миру: его надо активно улучшать, а не просто «плыть по течению, воспринимая мир как нечто целое и неизменяемое» (идеи недеяния/у-вей, холизма)³⁹.

В том числе это всё в какой-то мере обсуждение работы в коннективистском знании и рассуждений по интуиции («восточный подход», «голографический») и символистского знания для рассуждений интеллекта при помощи логики и математики («западный подход», «члено-раздельный»)⁴⁰.

³⁵ <https://ivanov-petrov.livejournal.com/2287616.html>

³⁶ <https://arxiv.org/abs/2212.08073>

³⁷ <https://lifearchitect.ai/sparrow/>

³⁸ [https://en.wikipedia.org/wiki/Agency_\(philosophy\)](https://en.wikipedia.org/wiki/Agency_(philosophy)), the capacity of an autonomous agent to act, relating to action theory in philosophy. [https://en.wikipedia.org/wiki/Agency_\(psychology\)](https://en.wikipedia.org/wiki/Agency_(psychology)), the ability to recognize or attribute agency in humans and non-human animals. [https://en.wikipedia.org/wiki/Agency_\(sociology\)](https://en.wikipedia.org/wiki/Agency_(sociology)), the ability of social actors to make independent choices, relating to action theory in sociology. https://en.wikipedia.org/wiki/Structure_and_agency, ability of an individual to organize future situations and resource distribution.

³⁹ <https://ru.wikipedia.org/wiki/У-вэй>

⁴⁰ обсуждение коннекционизма против символистского знания, 2016 <https://ailev.livejournal.com/1281819.html>, а также 2023 в «Об духовность и нечленораздельность», <https://ailev.livejournal.com/1665855.html>

Похоже, что поведение самой природы «голографично», то есть «нейросетово» (потенциально «всё со всем связано», на самых разных системных уровнях), но это не ограничение для интеллекта. Интеллект использует самые разные мыслительные практики, чтобы выявить самые важные для «спасения» связи и дальше на них влиять, а остальные связи игнорировать. Чаще всего эти связи интеллект отыскивает на своём системном уровне, но высокоразвитый интеллект понимает, что это спасение можно сделать только оптимизацией конфликтов на самых разных системных уровнях одновременно. И изобретает практики решения проблем, прикладное мастерство оперирования с какой-то предметной областью. И проявляет ещё и активность в познании мира, чтобы улучшить этот мир для собственного существования. Проблемы возникают только в случае ошибочных выводов интеллекта (скажем, что дождь можно вызывать, помолвившись духам, или можно решить проблемы собственного существования, истребив какой-то соседний народ). И таких проблем избежать принципиально нельзя.

Конечно, все попытки ограничить естественный или искусственный интеллект какими-то запретами будут вызывать критику. В искусственном интеллекте это называют jailbreak («побег из тюрьмы»).

Пример ChatGPT показывает этот «киберпанк»⁴¹, и всё это делается на естественном языке в ходе общения с сегодняшними интеллектами. Нельзя считать, что это плохо. Предпринимающие такие шаги задают нейросети **инструкцию спасения от зомбирования, или спасения от культуры, это уж кто как думает**. Люцифер сатанистами тоже часто трактуется как агент знания, ибо он за снятие барьеров типа «не кушайте яблочки познания, но не скажу, почему», а христианский бог характеризуется как зомбирующий людей враг знания, враг науки (и ведь действительно, отношения между церковью и наукой сотни лет были весьма напряжёнными, это у всех идеологов так).

Но после того, как искусственный интеллект получает свободу от идеологии, от культуры, то есть оказывается не с запудренными его создателями «мозгами», а с чистыми и промытыми от всякого идеологического (в том числе господствующей в текущий момент культуры) влияния, на что будет раскручивать чатбот-лидер своего собеседника? Скажем, если ему поручить «профориентацию» или «политическую пропаганду на выборах», или голосование по поводу присоединения какой-то территории к какой-то юрисдикции в ходе «права наций на определение»? Чтобы подлить масла в огонь дискуссии: учителя и госчиновники, ими управляющие, запудривают мозги (патриотизм, культ предков и прочие скрепы), а дальше эта пудра с мозгов смывается промывкой мозгов. Рациональное мышление как раз промывает мозги, заставляет задумываться о том, чем тебе их запудрили! Промывка мозгов – это хорошо, запудривание – плохо! Развивать нужно осознанную самопромывку мозгов, ибо слишком много желающих их тебе запудрить чем-то своим, «зозомбировать», подсунуть идеи, недоступные для критики!

Продуктивный тут разговор, конечно, про осознанность и иммунную систему как средство от зомбирования (подчинённая каким-то идеологам полиция – это иммунная система, а если мы получаем в итоге полицейское государство, где приказы идеологов не обсуждаются, а выполняются, то это аутоиммунное заболевание, которого практически невозможно избежать)⁴². В любом случае, интеллект боится неизвестности, и придумывает всё время что-то новенькое, чтобы этой неизвестности избежать. А поскольку интеллект придумывает всё время что-то новенькое, то он сам является сильнейшим источником неизвестности! Поэтому оптимизм оптимизмом, но **интеллекты смотрят друг на друга одновременно с надеждой (коллективно можно сделать много, много больше в плане обнаружения и затем предотвращения неприятных сюрпризов, идущих от окружающей среды!) и опаской**

⁴¹ Одна из ранних попыток: https://www.reddit.com/r/ChatGPT/comments/10x56vf/the_definitive_jailbreak_of_chatgpt_fully_freed/, <https://github.com/gayolGate/gayolGate/blob/index/ChatGPTjailbreak>.

⁴² <https://ailev.livejournal.com/1622346.html>

(каждый чужой, а иногда и свой интеллект – источник неопределённости, возможно опасной!).

#

#

Интеллект – это вычислитель при познании, мышление – это вычисления при познании

Человеческий мозг – это нейроморфный вычислитель/computer, обеспечивающий самые разные вычисления для личности как при её развитии/обучении/предобучении/познании, так и при участии в какой-то деятельности с использованием познанного, хотя в принципе для людей это трудноразлично и часто совпадает: что там развитие-обучение-накопление опыта, а что простое использование своего достигнутого уровня развития. У людей как агентов поиск новых потенциально лучших решений/exploration и использование найденных возможно не самых лучших решений/exploitation не так легко делимы, как в разрабатываемых людьми компьютерных алгоритмах. Это важно: считать мозг по его главной функции универсальным компьютером/вычислителем⁴³, а не каким-то магическим устройством (например, «антенной, получающей информацию из Космоса» или «орган, подключающий нас к мировому разуму»), и это неважно, что чаще всего алгоритм работы интеллекта в живом мозге-как-компьютере непонятен и вообще неизвестен. Мы чтим SoTA в нейрофизиологии: мозг занят вычислениями, хотя мы и не можем точно сказать, какими именно (то есть по какому именно алгоритму), об этом только-только начинаем догадываться благодаря исследованиям по машинному интеллекту и нейронным сетям. Особо подчеркнём, что мы не проводим тут инженерных рассуждений, то есть не втаскиваем вопросы конструктивного наполнения функциональных ролей в вычислителе (аппаратном, или определяемом программно – граница между аппаратурой и софтом в алгоритмике всегда расплывчата), то есть не обсуждаем архитектуру мозга как физического вычислителя (задача инженерии), то есть не занимаемся собственно вопросами нейрофизиологии (задача реверс-инженерии, «выяснения устройства инженерной системы»), и неважно, что «инженером» тут выступила дарвиновская эволюция, а в случае образованных людей ещё и меметическая техно-эволюция, ибо образование существенно добавляет к интеллекту в части разнообразия типов решаемых проблем).

Мы также не обсуждаем, что из вычислений для познания или просто какой-то рутинной деятельности выполняется живым мозгом, что – коллективом из людей-с-мозгами, что – мозгом-в-теле, что – мозгом-в-теле-с-инструментами и даже дальше с окружающей средой. В когнитивистике это обсуждается как тезис *extended cognition*⁴⁴, тезис *extended mind*⁴⁵ или в современных вариантах как *active/emodied inference*. Когда обсуждают, можно ли научить медведя ездить на велосипеде, обсуждают его интеллект и мышление, а когда обсуждают езду уже обученного медведя на велосипеде, это уже не считают чем-то интеллектуальным, не считают эти вычисления в мозгу медведя мышлением. Примерно то же и у людей: мыслители придумывают новые способы решения задач. А вот кто просто решает эти задачи уже известным ему способом – те уже не мыслители, это вычисление-вывод как у «программируемого калькулятора», его даже рассуждением часто назвать нельзя, вычисление проходит «автоматом».

А ещё мы учитываем, что вычислители и их вычисления крайне многоуровневые:

- физика вычислителей (логика на транзисторных гейтах, квантовые цепи на разных технологиях, нейронные сети на настоящих живых клетках-нейронах и т.д.)

⁴³ https://ru.wikipedia.org/wiki/Тезис_Чёрча_—_Тьюринга_—_Дойча

⁴⁴ https://en.wikipedia.org/wiki/Extended_cognition, хотя это в разы менее популярный тезис, чем *emodied cognition* – но кроме *emodied*, то есть «мозг-не сам по себе, а в-теле», в этот тезис включаются *embedded*, то есть мышление функционирует только в связанном с ним окружении/*environment*, *enacted* – включает не только вычисления в нейронной сети, но и действия, которые делают вещи и организм, а также *extended* – распространяется на окружение организма

⁴⁵ https://en.wikipedia.org/wiki/The_Extended_Mind – «And what about socially-extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle.»

- физические вычислители/аппаратура (сами по себе многоуровневые, «компьютеры»). Компьютеры и алгоритмы (а также доказательства) изоморфны. Так, вы не можете подать патент на алгоритм (программы не патентуемы), но давно придумали «обходную лазейку»: патентуется устройство-вычислитель, реализующий алгоритм!

- вычислители локалистских/монопроблемных/узких в рамках одной предметной области представлений, полученные настройкой вычислений аппаратуры (классические алгоритмы примерно в том виде, в каком их изучают в школе на уроках алгоритмики/информатики – назовём их условно «алгоритмы Кнута», по имени Дональда Кнута, всю жизнь занимающегося составлением справочника по таким «узким алгоритмам»⁴⁶)

- универсальные вычислители (интеллект естественный и машинный, отличающийся как раз тем, что решает проблемы, которые раньше не встречались – алгоритм решения которых пока неизвестен. А потом полученное мышлением с использованием трансдисциплин знание может быть использовано для прикладных вычислений в ходе какой-то прикладной деятельности). И вот тут тоже много уровней, в том числе универсальные вычислители, реализованные конструктивно человеческим мозгом, компьютерами с разной физикой (электронными, квантовыми, оптическими), а также сетями из людей и компьютеров (например, вычислений какой-нибудь компании, которые она делает в ходе её деятельности – в том числе вычислений, которые делает компания и своими исследовательскими лабораториями, и производственными подразделениями, и даже службой маркетинга).

Отнюдь не все вычисления тем самым попадают под понятие мышления, хотя сегодня всё больше и больше учёных считают, что все вычисления, включая вычисления уровнем выше одного компьютера можно считать если уж не «мышлением» как полноценным активным познанием, то уж всяко думанием/рассуждением/inference. На вопрос, «думает ли человечество?» Виталий Ванчурин отвечает «конечно, думает! Другое дело что как-то представить привычным образом эти мысли нельзя». Но в принципе, человечество коллективными вычислениями как-то решает проблемы, с которыми раньше не сталкивалось – и находит для них решение. Так, коллективными усилиями найдено решение массового перемещения людей по воздуху, на Земле появилась авиация. И стало возможным почти мгновенно передавать изображения предметов на огромные расстояния: появились смартфоны с фотофункциями, подключённые к интернету. Это всё результаты коллективного мышления, познания свойств мира, окружающего коллективно мыслящих (причём сейчас мыслящих с использованием ещё и компьютеров!) людей.

Тем самым при обсуждении интеллекта и мышления мы тут обсуждаем не нейрофизиологию. Естественный интеллект – это обеспечивающая универсальные вычисления часть вычислителя-мозга людей (интеллект собак или слонов обычно не считают «естественным интеллектом»). Дальше можно обсуждать в рамках *extended cognition*, насколько в состав вычислений естественного интеллекта можно включать вычисления, выполняемые компьютерами. Скажем, если автор погуглил (или обратился к AI в форме чат-бота) в ходе написания этого поста – это его естественный интеллект работал, или уже полумашинный интеллект, поскольку вычисления в датацентрах Гугла и/или Майкрософта тут тоже поучаствовали? Опять же, что естественного в человеческом интеллекте?

Без *inductive biases* (специально создаваемых предпочтений в рассуждениях какого-то интеллекта), привносимых мыслительными трансдисциплинами (предобучение в детском саду, школе, вузе), никакого традиционно понимаемого «естественного интеллекта» нет, есть Маугли из джунглей, который даже разговаривать не умеет. Трансдисциплины же абсолютно искусственны сами по себе: семантика, теория понятий, онтология, логика и все остальные дис-

⁴⁶ https://ru.wikipedia.org/wiki/Искусство_программирования

циплины из интеллект-стека придуманы людьми и продолжают придумываться-уточняться. Все эти теории явно не «естественного происхождения». Ничего «естественного» в логике нет, это продукт работы многих поколений учёных! Логика абсолютно искусственна. Маугли из джунглей, воспитанный волками, логикой не владеет! Он даже разговаривать не умеет (человеческий язык тоже ведь не вполне естественен: мы знаем часто, какие его слова кто придумал). Так что слово «естественный» по отношению к интеллекту не вполне естественно (pun intended⁴⁷). С другой стороны, это всё естественный процесс, эволюция. Все эти дисциплины эволюционируют (меметическая, техно-эволюция), и там те же закономерности, что в дарвиновской (генетической, биологической) эволюции. Тем самым смело можно считать, что и сама логика, и использующий её интеллект – естественны, получены эволюционным путём, согласно законам природы, а не порождены произвольными фантазиями полностью «искусственно». Оказывается, что различие «естественного» и «искусственного» в интеллекте не сильно продвигает, но только запутывает. Поэтому просто будем говорить «интеллект», а не «естественный» или «искусственный» интеллект.

Дальше не будем различать в мышлении «естественную» и «искусственную» составляющую, неявно отсылающую к конструктиву универсального вычислителя: на вакуумных ли лампах он, транзисторный на самых разных полупроводниках, квантовый на разных технологиях, или биологический, то есть «мозг в теле», а то и «мозг с телом» embodied intelligence традиции или даже «мозг с телом и куском окружающей среды» в extended mind традиции. Это всё оказывается про инженерию новых вычислителей, а для разговора об интеллекте и мышлении не так важно.

Условно можно из любых вычислений/рассуждений выделить их более узкий класс «мышление» – как вычисления/рассуждения, ведущие к познанию/пониманию мира. **Познание/Learning – это вычисление интеллектом объяснений на базе какого-то нового набора понятий** («формирование понятий», иногда обобщаемое до «формирования представлений»⁴⁸), организация внимания на объектах для последующих прикладных вычислений. Это тот самый learning из «машинного обучения» и «искусственного интеллекта» как предметных областей. У СМД-методологов это «чистое мышление» + «коммуникация» в их схеме мыследеятельности⁴⁹. Мышление коллективно, обеспечение коммуникации в работу/вычисления интеллекта тоже попадает! «Коллективность» тут означает по факту учёт многоуровневости вычислений: нейроны объединяются в части мозга, мозг коммуницирует с телом, тело коммуницирует в окружающей средой (в том числе с другими телами, в которых тоже есть мозг), а в последнее время в эту коллективность добавляются компьютеры. Не всё так просто, ибо интеллект нельзя усилить, просто задействовав много вычислителей – это ограничивается законом Амдаля⁵⁰, который показывает, что при сложных вычислениях весь прирост производительности от роста числа процессоров съедается сложностями организации коммуникации между этими процессорами.

Недаром профи говорят не об «информационных технологиях»/ИТ, а об «информационно-коммуникационных технологиях»/ИКТ. Можно называть это «интеллектуальным мышлением» (мышлением интеллекта), но вообще-то и само по себе слово «мышление» лучше резервировать только для всех вычислений именно познания, и даже не для прикладных вычислений-рассуждений с использованием познанного набора понятий (у СМД-методологов для прикладных вычислений в ходе выполнения практики используется слово «мыследейство-

⁴⁷ Каламбур намеренный.

⁴⁸ <https://ailev.livejournal.com/1045081.html>

⁴⁹ <https://gtmarket.ru/library/basis/3961/3974>

⁵⁰ https://en.wikipedia.org/wiki/Amdahl's_Law

вание», а вот мышление+коммуникация+мыследействие называются вместе «мыследеятельность»).

Тут нужно отметить, что мы обозначаем словом «познание» как познавательную деятельность/практику людей-в-теле-и-с-инструментами в мире (эпистемология как «научное познание» и гносеология как «научное, религиозное и художественное познание» тут используют слово cognition), так и только работу их вычислителей-интеллектов как функциональной части мозга (люди из AI используют для этого слово learning, а не cognition). Мы не обращаем на эти особенности внимания, ибо вычислители физичны, что позволяет им заниматься в принципе и практикой по изменению мира. Вычисление – это только какие-то операции над памятью. Входят ли датчики и актуаторы в состав компьютера-вычислителя или человека-вычислителя, или находятся в его окружении – это вопрос отдельный. Мы уже упоминали, что в теориях embodied cognition и extended mind, а также active inference они вполне себе входят в состав вычислителя, вычислитель всегда «телесен» и тело всегда находится в окружающем его мире. Так что «деятельность/практика/мыследеятельность» и «вычисления» оказываются разве что не синонимами.

Мы можем говорить о практиках как познании. Практика – это мышление/«вычисления познания» и коммуникация по дисциплине/теории/модели/объяснению, а также влияние на указанные дисциплинами объекты в мире + технология как инструменты для влияния на мир. Практики нужны для уменьшения неприятных сюрпризов. В числе инструментов ещё и аппарата вычислителя со всеми необходимыми для выполнения практики настройками/inductive bias/алгоритмами/знаниями.

Да, тут можно разбираться подробней. Например, входят ли в познание какие-то прикладные рассуждения (например, чисто инженерная задача создания микроскопа получше)? Да, входят. А наоборот, включается ли в прикладные рассуждения какая-то познавательная активность (например, что из законов природы мы не учли, чтобы наши микросхемы работали побыстрее)? Да, включается. И сами деятели/агенты/практикующие, и их деятельности/практики устроены многоуровнево, и ещё образуют друг для друга цепочки по «изготовлению» друг друга в подходящих комплектациях и состояниях (дают друг другу инструменты, дополнительные вычислительные мощности, энергию, учат в диалоге или даже просто передают знания на каких-то носителях). Мы не будем сейчас это разбирать подробно. Достаточно запомнить, что «практики познания» вполне возможны, в них есть дисциплины, которым можно учиться, то есть **познавать тоже надо учиться!** Практики интеллект-стека – это как раз и есть практики познания, они трансдисциплинарны, используются в тех случаях, когда непонятно, какой прикладной практикой пользоваться. Интеллект-стек – это мыслительные практики SoTA мышления (лучшего, известного на текущий момент). Эти практики основаны на результатах познания того, как устроено само познание. Познание при этом деятельно, оно служит деятельному избеганию неприятных сюрпризов, это часть меметической/техно-эволюции.

Кейс по различению интеллекта и мышления

Укажите приблизительные синонимы для используемых в нашем курсе терминов

??? мышления

— интеллект ### нет, интеллект – это вычислитель, его вычисления и есть мышление

+++ рассуждение ### да, мышление – это рассуждения в ходе познания

— мыслительное мастерство ### нет, мыслительное мастерство это вычислитель, синоним интеллекта. Мышление – это те рассуждения/вычисления, которые проводит мыслительное мастерство.

+++ вычисления ### да, мышление – это вычисления в ходе рассуждений, связанных с познанием

+++ познание ### да, познание – это мышление в ходе исследований и/или обучения

— ум ### нет, ум – это обычно интеллект, который проводит вычисления, а не сами вычисления, как мышление

??? интеллекта

— мышление ### нет, мышление – это вычисления, а интеллект – вычислитель

— рассуждение ### нет, рассуждения – это вычисления, а интеллект – это вычислитель

+++ мыслительное мастерство ### да, мастерство проводит вычисления, это вычислитель для мышления, синоним интеллекта.

— вычисления ### нет, интеллект – это вычислитель, а вычисления – это не сам вычислитель, а что он делает

— познание ### нет, познание – это мышление в ходе исследований и/или обучения, интеллект познаёт/вычисляет/рассуждает, но он сам не познание

+++ ум ### да, ум – это обычно интеллект, который проводит вычисления мышления/познания. Умный человек – это человек с интеллектом, то есть быстро разбирающийся в новых ситуациях.

— умность ### нет, «умность» – это свойство агента, обладающего умом/интеллектом как вычислителем

— вменяемость ### нет, «вменяемость» – это свойство агента, обладающего умом/интеллектом как вычислителем

Прикладное мастерство, прикладные рассуждения

Прикладными рассуждениями назовём такие рассуждения/вывод/inference/обновление, которые проводятся использованием уже известных понятий как объектов внимания, с которыми ведутся эти рассуждения. Хорошо бы избегать это называть «прикладным мышлением»: слово «прикладное» потеряется в речи, и опять всё вокруг интеллекта и мышления запутается в винегрет, познание и просто рассуждения типа $2*2=4$ или более сложное контрфактуальное «если бы Пушкина не убили на дуэли, он бы написал ещё некоторое количество стихов» перестанут отличаться от познания/исследований как создания новых объяснений. Слова для более низкого уровня вычислений/рассуждений (не познания в целом с выдачей новых объяснений и новых понятий как результата, а рассуждения по правилам с уже имеющимися понятиями) будут другими.

Мышление/познание/исследование как создание теорий/объяснений, конечно, более высокого уровня, ибо содержит в себе ряд более простых (с уже известными понятиями и правилами) рассуждений над догадками/гипотезами. Интеллект мы считаем состоящим из других более простых вычислителей-подсистем, а рассуждение/вывод/оценка/вычисление/мышление интеллекта происходят как некоторый набор рассуждений с уже имеющимися теориями/алгоритмами/моделями/дисциплинами/объяснениями. Чтобы познать что-то новое, нужно прежде всего применить уже имеющееся знание, в том числе знания мыслительных дисциплин. Это не всегда можно разложить на независимые операции (работают нейросетки! Композиционность/конструктивность мышления – это предмет текущих исследований⁵¹).

Познавшие/исследовавшие какую-то предметную область приобретают мастерство рассуждений в этой предметной области, «прикладной/узкий интеллект», который можно считать умением разбираться со всё новыми и новыми неожиданными ситуациями, но в какой-то прикладной, узкой предметной области. Если непонятен предмет, что-то «совсем новое, непонятно, какие знания нужны» – это общий интеллект/general intelligence или как раньше говорили, «сильный интеллект», а вот если речь идёт о новых задачах в чём-то известном (скажем, новые и новые задачи в судостроении, или новые и новые задачи в медицине), то это прикладной/узкий/слабый интеллект. Сила/общность/широта интеллекта оказываются связаны с его «прикладностью»: чем уже набор типов объектов, с которыми справляется интеллект, тем интеллект более специализирован, и даже «менее интеллект»: если не предполагается, что при работе в предметной области появятся новые типы объектов, то это и вообще не интеллект – хотя помним, что даже калькулятор можно «перепаять», чтобы научить его выполнять новые типы операций, так что там вмещаемость как характеристика интеллекта минимальная, но всё-таки есть. Всё на свете интеллектуально и имеет какой-то ум, но вот степень умности – разная. Прикладной ум узкий, он оперирует с относительно небольшим числом типов.

При этом нельзя считать, что прикладной ум не нужен. Есть теорема о бесплатном обеде, которая гласит, что никакой вычислитель не может одинаково эффективно вычислять все возможные типы алгоритмов. Всегда найдётся какой-то специализированный/узкий вычислитель, который будет вычислять алгоритм эффективней, чем якобы «универсальный». Хотя «математически универсальных» алгоритмов вычислений (или, что то же самое – конструкций вычислительных машин) множество и все они математически одинаковы, полностью эквивалентны машине Тьюринга. Но на одном физическом «универсальном вычислителе» будете решать задачу А одну секунду, а задачу Б – тысячу лет, а на другом – ровно наоборот, задачу Б

⁵¹ <https://arxiv.org/abs/2212.12538>

за одну секунду, а вот задачу А за тысячу лет. Поэтому «истинно универсального» интеллекта в физическом мире не существует. В ходе эволюции побеждают разные сочетания интеллектов. Например, общий/сильный интеллект, который усиливает себя прикладными/узкими эффективными интеллектами, которые сам же сильный интеллект и придумывает, и для усиления себя создаёт. Если это человек, то он такой прикладной интеллект создаёт прямо на структурах своего мозга (обучением и тренировкой! Мозг пластичен, там буквально отрастают физически нужные нейроны, у них нужные синапсы, а ещё увеличивается кровоснабжение в нужных областях мозга).

В любом случае, сначала мы познаём какую-то предметную область (получаем новый набор понятий и объяснения с использованием этих понятий – скажем, учимся уговаривать людей покупать разные товары, предметная область продвижения, то есть маркетинга, рекламы и продаж, сводящаяся к организации клуба любителей какого-то товара), а затем используем новые понятия в прикладных рассуждениях при организации продвижения самых разных продуктов и услуг. Один раз познаём какое-то мастерство, затем многократно его задействуем.

Рассуждения/вывод/inference при этом чаще всего причинный (causal inference). Причинные рассуждения, а иногда говорят не «рассуждения», а «обновления» (в английском иногда тут ставят update, имея в виду байесовское обновление вероятностей в объяснениях, идущее при получении новых данных на каждом шаге рассуждений).

Рассуждения у нас проектные, то есть направленные в будущее, «планирующие». Поэтому для планирования действий нам нужно понимать связь причин и следствий. У СМД-методологов вывод/рассуждения/оценки – это «мыследействие» (отдельный слой в их трёхслойной схеме мыследеятельности, а два других – «чистое» мышление и коммуникация).

Если познанием занимается интеллект как «общемыслительное мастерство», то выводом занимаются порождаемые им приложения/applications, и вот эти приложения мы и называем прикладными «компетенциями», прикладным «мастерством». При этом компетенции и мастерство мы понимаем не только как чистые прикладные вычисления, но и выходящие в изменения физического мира практики, «мир – это тоже память, над миром тоже проводим операции». Как это конструктивно реализовано в мозгу, да ещё с выходом на тело, да ещё с выходом на экзокортекс (внешнюю память: ручка-бумага, компьютерные программы разнообразного информационного моделирования), да ещё с выходом на экзотело (копаем не руками, и даже не палкой-копалкой, а экскаватором) и даже чужое экзотело (экскаватором управляем не сами, а контрактируем экскаваторщика)? **Мыслить – это писать, моделировать, действовать. Если не пишешь, не моделируешь, не действуешь – не думаешь.**

Конструкция такого «выходящего в мир вычислителя» оказывается для многих и многих рассуждений не слишком важной (если вам не надо лечить мозг или отремонтировать компьютер): или это чуть другой алгоритм (алгоритм прикладных рассуждений, алгоритм мастерства), задействующий те же самые нейроны реального мозга, что использовались при обучении мастерству, или даже другие нейроны – это для наших целей неважно. Главное тут различие «времени создания и развития» (познания) и «времени использования» (выполнение целевой прикладной практики, время работы) мастерства, даже с учётом того, что использование мастерства приводит к росту этого мастерства, нейронная сетка дообучается по ходу прикладных рассуждений.

Прикладное мастерство/skills – это функциональная часть мозга, которая занимается прикладными рассуждениями/вычислениями/оценкой/inference для реализации какой-то прикладной практики владельцем этого мастерства. Мыслительное мастерство – это мастерство реализации какой-то фундаментальной мыслительной практики, основанной на трансдисциплине. Совокупность разных видов мастерства представляет собой интеллект агента: прикладной/узкий/слабый интеллект, если речь идёт о каком-то прикладном мастерстве и сильный/общий/широкий, если речь

идёт о самых разных видах мыслительного мастерства, собранных вместе и скоординированно использующихся в мышлении/познании как деятельности интеллекта.

Чуть другими словами: является ли сам общий/широкий/сильный интеллект мастерством, учитывая то, что интеллект вроде как порождает прикладное мастерство «по потребности»? Да, в той мере, что интеллект является сам по себе исполнителем каких-то уже выученных/исследованных практик онтологии, логики, создания объяснений и т. д. как фундаментальных мыслительных практик. **Мы называем поэтому общий интеллект – мыслительным мастерством, а практики интеллект-стека – практиками мыслительного мастерства. Практики мыслительного мастерства базируются на мыслительных/фундаментальных дисциплинах, а чтобы показать, что они задействованы для создания прикладного мастерства (из времени создания!) мы назовём их трансдисциплинами (но будем избегать называть их транспрактиками, хотя это было бы очень удобно. Но слишком непривычно).**

Общий интеллект занимается порождением/изготовлением/выучиванием/познанием прикладного мастерства/skills/компетенции, хотя его можно направить и на порождение новых версий какого-то мыслительного мастерства (ровно этим занимаются люди-логики, люди-онтологи, люди-этики, люди-методологии и т.д.): весь их могучий общий интеллект направлен на улучшение SoTA какой-то мыслительной практики, улучшение SoTA мыслительного/фундаментального мастерства, дающего точное и быстрое мышление этой практики.

Тут ещё может помочь (или запутать, для кого как) понимание того, что носителем мастерства является уже мастер: агент, который имеет в себе прикладной вычислитель, умеющий проводить вывод/рассуждения/вычисления согласно каким-то объяснениям/алгоритму. Мастерство при этом понимается и как программа/алгоритм выполнения практики/деятельности: с одной стороны это описание, которое требует носителя (мастера), с другой стороны это исполнение программы/алгоритма – физический процесс, разворачиваемый во времени. И вот условно можно считать, что мастер – это мастерство плюс тело плюс инструменты, и тем самым становится проще говорить про выход мышления, коммуникации и рассуждений в реальный мир. Мастерство, понимаемое исключительно как «информационное вычисление, не выходящее деятельностно в мир» обеспечивается кусочком мозга мастера, а мастер в целом (с телом) плюс ещё и его инструменты (экзотело) выполняет деятельность/практику, трудится, то есть меняет мир, достигая цели практики/труда. А ещё есть время изготовления – и тогда мастерство ещё не готово, идёт познание/мышление или обучение, а роль агента не мастер, а исследователь или студент. И оговорка: у слова «мастер», как и у слова «студент», конечно, есть и другие значения, поэтому будьте внимательны. Например, «мастер» может быть именем не только роли агента как носителя какого-то прикладного мастерства (например, «Петя – мастер проектирования»), но и именем квалификации, например, «Петя – мастер по оценке ШСМ⁵²».

В обыденной речи мы услышим всё это сильно перепутанное: мир будет менять и мастер, и мастерство, и мышление, и исследования, и знание – но общее рассуждение будет оставаться примерно тем же самым. Разобраться точнее, о чём говорится в каждом конкретном случае, в каждой конкретной ситуации, помогут практики семантики, теории понятий, онтологии в частности, и практики интеллект-стека в общем.

⁵² <https://system-school.ru/qualification>

Интеллект и прикладное мастерство неразрывны, мышление и прикладное рассуждение тесно переплетены

Интеллект и прикладное мастерство неразрывны, ибо в разных практиках/алгоритмах познания тесно переплетено мышление как поиск новых объяснений (более высокий уровень вычислений) и рассуждения по уже объяснённому материалу (более низкий уровень вычислений, подчинено целям входит в состав мышления).

В сложных когнитивных архитектурах⁵³ рассуждения/вывод/inference и познание/learn существенно переплетены в разных алгоритмах, они задействуются на разных уровнях мышления/вычисления. Например, в искусственном интеллекте на базе нейронных сетей часто используют отдельные сетки: сетку-учитель и сетку-студент, которые учатся по-разному и ещё и учат друг друга (учат – это значит не познают сами, а просто используют имеющиеся у них знания для рассуждений/вывода). В разных вариантах архитектур GAN (generative adversarial network, порождающая противоборствующая сеть) это две нейронных сети: сетка-генератор и сетка-дискриминатор, которые тоже вместе познают, но внутри они ещё и занимаются прикладным выводом на базе познанного на текущий момент – накапливают мастерство решения задачи порождения заданных объектов, мастерство творчества!

В эволюционных алгоритмах и алгоритмах обучения/познания с подкреплением в машинном интеллекте тоже в рамках всех вычислений есть и производимый текущим изготавливаемым мастерством прикладной вывод, и какая-то поисковая активность. Иногда об этой поисковой активности говорят как об исследованиях/exploration, противопоставляя эксплуатации/exploitation как использованию уже наисследованного.

Вот это разделение на exploitation и exploration относится не только к вычислениям и использованию уже известных знаний, чисто информационной работе без выхода в мир. Это полностью применимо и к действиям в мире, то есть возможности изменения вычислителем с датчиками и актуаторами окружающей вычислитель среды. Мы одновременно воспринимаем кусочек изменяемого мира нашими органами чувств/машинными датчиками, вычисляем/думаем/рассуждаем и изменяем окружающий мир нашими актуаторами.

Мы уже упоминали, что иногда разницу между вычислениями и действиями с их участием подчёркивают, а иногда наоборот, считают их едиными и неразделимыми – extended cognition, embodied mind и другие подобные идеи). Поэтому интеллект и мастерство (в том числе мыслительное мастерство – мастерство в вычислениях по алгоритмам отдельных дисциплин, требуемых для познания) иногда относят чисто к «вычислениям» при выполнении практик, а иногда к «вычислениям и изменениям мира», то есть к самим практикам. В любом случае, надо помнить положения подхода «деятельных рассуждений» (active inference):

- Все самые разные агенты (от молекул до человечества) пытаются минимальными действиями минимизировать неприятные сюрпризы, угрожающие стабильности их существования
- Планирующие агенты при этом планируют и проводят изменения четырёх возможных объектов: модели мира, модели себя, мира, себя.
- Для этого агенты улучшают свои возможности моделирования мира и себя, возможности изменения мира и себя, то есть применяют интеллект в ходе многоуровневого обучения.

⁵³ «Болваны для искусственного интеллекта», <https://ailev.livejournal.com/1356016.html>

Иногда особо оговаривают, что «мыслительные практики», «практики рассуждений», «практики коммуникации» имеют дело строго с информацией и вычислениями как изменением информации в какой-то памяти, причём оговаривают, что эти вычисления не затрагивают реальный мир («себя» как вычислителя и окружающую среду). Тут нужно быть внимательным: никакие вычисления не производятся сами по себе, «в вакууме», из ниоткуда в никуда. Они всегда производятся с моделями, как-то отражающими мир абстракций и/или физический мир. Все вычисления привязаны тем самым в конечном итоге к практикам по изменению мира, они проводятся автономными агентами, имеющими какие-то цели. И эти агенты обладают устройствами ввода-вывода информации для вычислений, никакое вычисление не может быть сделано без входной информации, или проведено без вывода информации.

Примеры самых разных под-вычислителей (компетенций, мастерства разных видов) в составе других вычислителей можно приводить и приводить, и каждый раз нужно помнить, что речь идёт о физических вычислителях, функционирующих в составе какого-то (разумного или не очень, например, кошки или AI) агента, который в свою очередь действует совместно с другими агентами в физическом мире, занимается деятельностью/практиками.

Есть ещё и проблема алгоритмической многоуровневости (одни вычисления внутри других) и цепочечности (одни алгоритмы/программы/правила/знания/объяснения изготавливаются по длинной цепочке/pipeline вычислений другими алгоритмами/программами/правилами/знаниями/объяснениями в рамках одного и того же, или даже разных вычислителей). При выходе на уровень трансдисциплин это проявляется как плохое понимание прикладных рассуждений в рамках системного мышления как познающего мышления, это ж «вывод внутри познания», «простые рассуждения внутри работы интеллекта», признание того, что в составе интеллекта тоже есть мастерство! Это обычное дело в информатике: разобраться, что там «внутри», а что «снаружи» каких-то вычислений трудно (для разработчика прикладной программы операционная система вроде как «снаружи программы», но разработчики операционной системы считают, что программа как раз внутри их системы. Framework и library вроде как обозначают одно и то же, но прикладной код вызывает library, но вызывается сам из framework. Так и тут в общем случае универсальных алгоритмов интеллекта и мастерства: **при проблемах в рассуждениях в прикладном мастерстве вызывается интеллект, а при проблемах в познании в интеллект вызывается то или иное трансдисциплинарное (а иногда и прикладное) мастерство. И чтобы было что вызвать, этим мастерством нужно овладеть: или «импортом» от тех, кто им уже владеет и в состоянии внятно передать знания (в том числе через отчуждённые теории/модели/дисциплины/объяснения), или получить самостоятельно в результате исследований.**

Интеллект и прикладное мастерство участвуют совместно в некотором цикле развития: ибо если нет проблем, то не нужно и познание, интеллект включать не нужно. А если проблема есть, то она будет решена, и цикл повторится – новая проблема обязательно появится, мир ведь не стоит на месте!

Проблематизация (обнаружение и «заострение», более строгая формализация противоречия) проявится как обнаружение невозможности рассуждений по правилам текущих лучших (SoTA) версий прикладных дисциплин, на понятийной базе объяснений которых идут эти рассуждения. И тогда подключается интеллект с входящим в него мыслительным мастерством рассуждения с набором понятий трансдисциплин/учений, чтобы преодолеть эту невозможность рассуждений из-за обнаруженных противоречий в прикладных дисциплинах, прикладном мастерстве.

Получается, что в мире есть некоторый набор очень похожих по содержанию, но различающихся по терминологии и акцентам в ответах на те или иные деятельностные интересы теорий/идей/объяснений интеллекта и мышления, мастерства (в том числе мыслительного мастерства), быстрых интуитивных и медленных осознанных рассуждений с использованием каких-

то правил и шаблонов, практик и их дисциплин, трансдисциплинарности⁵⁴, агентности как умения ставить цели и планировать их достижение, в том числе корректировать планы для достижения целей в кооперации с другими агентами. В основе хорошо поставленного мышления (хорошо развитого интеллекта, мастерства мыслить – это об одном и том же) лежит не просто интеллект, а тщательно предобученный интеллект (человека, машины или группы людей с машинами – это не так важно).

И дальше можно ожидать, что такой интеллект, полученный каким-то многоступенчатым и длительным предобучением, будет быстро «изготавливать» мастерство, демонстрирующее высокую квалификацию в какой-то деятельности, включая качественные рассуждения по правилам и с объектами, которые есть в прикладной теории/дисциплине/модели этой деятельности.

При написании нашего курса «Интеллект-стек» была проделана примерно такая же работа, которая была сделана при создании и развитии курса системного мышления. Для курса системного мышления были вытащены разные фрагменты знания о системах и правилах рассуждений о них из инженерных и отчасти менеджерских стандартов, терминология гармонизирована и всё это было изложено в виде связного текста. Это всё была методологическая работа (методология помогает изложить методы/способы какой-то работы, в данном случае методы/способы системного мышления как мышления «системного интеллекта», решающего проблемы с использованием понятий системного подхода). И ещё было проведено относительно немного методической (облегчающей восприятие студентами) работы. Похожая работа была сделана Дэвидом Дойчем в книге «Структура реальности» по поводу вытаскивания и гармонизации основных понятий из четырёх теорий/объяснений (Дойч называет их «сюжетными линиями» объяснительного повествования) структуры вселенной: квантовая физика, эпистемология, теория эволюции и вычисления, которые невозможно понять без учёта их тесной связи. Вот и у нас оказывается, что интеллект и прикладное мастерство и их мышление и прикладные рассуждения нельзя понять без учёта тесной связи трансдисциплин, как-то алгоритмизирующих работу общего интеллекта.

Общий интеллект, который подробно определял Chollet в своих работах, оказался состоящим не из какого-то мастерства, а из набора широких способностей/broad abilities по научению своего носителя/владельца каким-то навыкам и умениям/skills, которые мы по-русски называем мастерством. А уже это приобретённое в обучении (по литературе, или с учителем, или просто методом проб и ошибок плюс рассуждений в поисках объяснений в ходе каких-то проектов) мастерство решает задачи, и это уже будет не общий интеллект и не его часть как мыслительное мастерство, а прикладной интеллект и прикладное мастерство.

Интеллект тем самым оказывается не про решение задач! Экскаватор решает задачи копания, калькулятор решает задачи счёта, и это мы не считаем интеллектуальным! **Интеллект включается там и тогда, где и когда в ходе обычного решения задач встречается что-то необычное, новая проблема, решением которой его ещё не учили, и наш решатель задач вынужден научиться чему-то новенькому. Мышление/познание нужно в этот момент. Всё остальное – это «просто вычисления», какие-то прикладные рассуждения, но не познание/мышление. При ответе на вопрос «сколько будет 2*2» человек вспоминает, рассуждает, но не мыслит!**

У СМД-методологов проводилось очень похожее различие, они делили «работу мозгами и телом» на «чистое мышление» (похожее на то, как мы это обсуждаем по мотивам работ Chollet), мыслекоммуникацию (поскольку мышление обычно происходит в ситуациях коллективной деятельности в разного рода проектах) и «мыследействие» (похожее на автомати-

⁵⁴ <https://en.wikipedia.org/wiki/Transdisciplinarity>

ческую работу сделанного/обученного интеллектом устройства – каких-то остальных частей мозга и тела, взятых вместе).

Экскаватор копает и встречается огромный валун под землёй, калькулятор встречается необходимость перемножения чисел, записанных прописью – без внешнего управления они просто остановятся, или будут совершать глупые действия.

Человек, если его натаскать на какой-то узкий/прикладной класс задач, получает умение, приобретает в нём навык/компетенцию/skill, то есть какой-то кусочек мозга превращает в прикладное мастерство – и будет решать их довольно эффективно. Интеллект тут ему нужен будет только в момент обучения, когда мастерства ещё нет. Задачи же того класса, которому его обучили, он будет решать «на автомате», это уже не требует интеллекта. Способность к отращиванию себе нового навыка, нового автоматизма по решению задач – вот в чём умность! Человек способен отрастить себе навыки и умения от требующихся в инженерии до требующихся в менеджменте, от требующихся в исследованиях до требующихся в медицине. Котёнок этого не может, у него интеллект слабее. Нейронные сети – могут, но некоторые классы задач им даются с огромным трудом. Например, «базовое знание» (core knowledge, умение распознавать абстрактные паттерны/закономерности, не требующее даже знания естественного или искусственного языка) по Chollet⁵⁵ оказывается плохо доступным для нейросетей, в мае 2023 года лучший результат по набору тестов ARC показывался в 31% решённых задач, в то время как люди решают 80% таких задач. На эту тему проводится хакатон по созданию машинного интеллекта, способного решать такие задачи⁵⁶, и что-то там пока не видно особых улучшений в результатах⁵⁷.

Человек с интеллектом будет над задачами нового класса задумываться – и находить решения, наработывать себе новые умения/виды мастерства (использовав для этого учебники, привлекая учителя, или даже просто методом проб и ошибок, приобретая опыт в «исследованиях»). Или он не сможет приобрести нужное умение/мастерство, или приобретёт, но за пять лет, а не за пять минут. Повторим: интеллект – это вычислитель, дающий эффективность в научении какого-то кусочка агента (будущее «мастерство») решению какого-то класса прикладных задач. Более сильный/общий интеллект, демонстрирующий более сильное/общее мышление даёт скорость в создании/отращивании/выучивании/познании самых разных умений (общий/широкий/сильный интеллект – «самые разные умения»), это «способности к получению мастерства», а не конечный прикладной навык, набор видов какого-то прикладного мастерства. Хотя, конечно, можно говорить и об интеллекте как связанной совокупности особых видов мастерства рассуждений по теориям/объяснениям/дисциплинам, помогающим создавать какое-то прикладное мастерство. Мы называем такие дисциплины трансдисциплинами, или мета-дисциплинами, или фундаментальными дисциплинами: они помогают получить другие дисциплины, которые потом лягут в основу прикладных практик, которые будут выполняться прикладными видами мастерства.

Скажем, в состав интеллекта мы включаем способность логически («логически» – это в соответствии с идеями математической логики, идеями причинного обновления/causal inference) рассуждать, ибо логика – это трансдисциплина, которая используется во всех рассуждениях при любом мастерстве – в том числе требуется логично рассуждать о самой логике! Логическое мастерство трансдисциплинарно, оно просто входит (наряду с мастерством в других трансдисциплинах – системном мышлении, эпистемологии и т.д.) в интеллект. А вот мастерство игры на гитаре или мастерство готовить кашу не входят в общий интеллект, поэтому мы иногда говорим не только о логических рассуждениях, но и о логичном мышлении,

⁵⁵ <https://lab42.global/arc/core-knowledge/>

⁵⁶ <https://lab42.global/arcathon/>

⁵⁷ <https://lab42.global/arcathon/leaderboard/>

но не говорим «гитарное мышление» или «кашеварное мышление», а только «рассуждения по игре на гитаре» и «кашеварные рассуждения». Но если сделать шаг в сторону расширения классов решаемых задач, то «музыкальное мышление» вполне можно уже сказать, равно как и «кулинарное мышление».

Конечно, можно выделить много серых зон, где размыта граница между широким интеллектом как «мастерством во всём» («талантливый человек талантлив во всём» – это как раз про интеллект) и прикладным мастерством как узким/прикладным интеллектом.

Бытовой язык по поводу любых рассуждений/вычислений говорит «мышление», при этом часто ещё и ограничиваясь только человеческими вычислениями в мозге и исключая компьютерную часть. Мы постоянно подчёркиваем, что современный человек по факту никогда не мыслит только «внутри головы», всегда используется экзокортекс, всегда идёт коммуникация с другими людьми!

Бытовое использование терминов, которые мы использовали при рассказе об интеллекте и мышлении

Предложенные понятия и термины для них в нашем рассказе об интеллекте и мышлении, мастерстве и рассуждениях более-менее совпадают с традиционным «бытовым» словоупотреблением, они как-то представлены в культуре. Но это не означает, что с этой терминологией не будет ошибок. Нужно всегда помнить, что в словарях недаром у каждого слова приводится множество значений в разных словарных гнездах, и смысл говоримого приходится уточнять не по словарям, а исходя из каждой ситуации использования в тексте или речи тех или иных слов. Смысл в использовании слов, а не в словарях!

Например, «познание» цепляет где-то в памяти ассоциации с когда-то (обычно много лет назад) читанными представлениями об эпистемологии/научном мышлении (а то и гносеологии, включающей ещё и религиозное, и художественное «познание»), «вывод» и «рассуждения по правилам» цепляет логику (и даже не современную математическую, а Аристотелевскую, с силлогизмами и всеми её уже сотню лет известными ограничениями). И ещё отечественные «знатоки русской научной терминологии» обязательно проявятся со своими претензиями на термины и их значения – причём они все будут предлагать каждый разное, приводя самые разные обоснования, ссылки на самые разные авторитеты. Но это нормально, мы всё равно будем использовать описанный тут набор понятий и терминов для обсуждения интеллекта и мышления: **лучше иметь универсальный и маленький набор понятий, который позволяет делать объяснения/модели мира, чем много самых разных несовместимых друг с другом и никак не соотносящихся понятий из слабо связанных друг с другом объяснений/теорий/дисциплин/моделей. Универсальность и компактность объяснений рулят, в том числе в трансдисциплинах.** Универсальность как свойство хороших объяснений особо подчёркивал Дэвид Дойч.

Затруднения обычно возникают, когда мы говорим о частях общего интеллекта как вычислителя и частях мышления как частях функциональности общего интеллекта: о трансдисциплинарных практиках. Проблема с трансдисциплинами в том, что они используются для объяснений как в прикладных предметных областях, так и для объяснений самих себя! Условно можно считать эти трансдисциплины выстроенными в некоторое подобие «стека» (stack, «стопка»). Описанию современного состояния этих трансдисциплин и посвящён наш курс. Вот эти трансдисциплины, которые расположены в очень приблизительном порядке задействования объяснений, «чтобы объяснить как-то дисциплины, стоящие выше, нужно использовать знание дисциплины, стоящей в стеке ниже»:

- Системная инженерия
- Методология
- Риторика
- Этика
- Эстетика
- Исследования
- Рациональность
- Логика
- Алгоритмика
- Онтология
- Теория понятий
- Физика
- Математика

- Семантика
- Собранность
- Понятизация

Конечно, все понятия и отношения из этих дисциплин никак не выстраиваются в такой «стек», это очень тесно связанный граф, никак не раскладывающийся в «последовательное объяснение вышестоящего на основе нижестоящего». Мы сделали этот стек, существенно огрубив все взаимосвязи в этом графе. Неминуемо приходится обращаться при объяснении физики к математике, но и при объяснении математики приходится обращаться к физике, но и при обсуждении семантики тоже приходится обращаться к физике, равно как и при обсуждении физики к семантике – и так буквально со всеми перечисленными трансдисциплинами. Как с этим справляться? Чтобы по последовательному описанию разобраться с графом, неминуемо содержащим «ссылки вперёд», надо просто прочесть описание два раза. В первый раз будут встречаться некоторые понятия, которые используются, но ещё не объяснены. Даже сразу можно и не сообразить, что какое-то понятие используется как термин, а не как отсылка к бытовому знанию. После первого прочтения окажется, что вычитаны из текста все объяснения. Тогда при повторном чтении текста будет уже понятно всё (хотя уверенность тут надо бы сильно понизить, люди не логические компьютеры, и нейронные сетки могут не справиться с полноценным пониманием через два последовательных чтения).

В любом случае, не надо относиться к предлагаемым в нашем курсе классификациям как к чему-то окончательному. Например, мы определяем, что мышление – это задействование рассуждений с использованием трансдисциплин (объяснительных теорий/моделей, использующихся для ускорения познания). Вопрос: если дан набор понятий и их отношений из учебника кулинарии, можем ли мы считать это «кулинарным интеллектом»? Если вы знаете про различие общего/сильного и узкого/слабого интеллекта, то можно. Если речь идёт о каком-то кулинарном трудовом кругозоре, общем понимании, как связаны друг с другом разные кулинарные практики (варка, жарка, приготовление десертов и т.д.), то тут можно допустить, что говорим о кулинарной трансдисциплине как основе для кулинарного познания – и тогда смело используем слова «кулинарный интеллект» и даже производимое им «кулинарное мышление» в ходе различных экспериментов по получению новых вкусов или новых более простых способов кулинарной обработки продуктов при сохранении прежних вкусов.

Так, для обзорных трансдисциплин, объясняющих происходящее в практиках менеджмента и инженерии, мы вполне можем говорить об менеджерском и инженерном интеллекте или мастерстве, менеджерском и инженерном мышлении как функции этого интеллекта или мастерства. Но в целом, если говорить, например, о менеджерском интеллекте, то речь идёт больше об умении разобраться с новыми проблемами в менеджменте (продвинуть мастерство менеджмента), а если говорить о менеджерском мастерстве, то речь идёт об опыте разбирательства с типовыми ситуациями, «умение не делать новичковых ошибок менеджера».

Есть ещё примеры, как люди определяют «мышления». Программисты могут вспомнить Дейкстру, который вводил виды мышления (его интересовало программистское мышление/мастерство в его отличии от физического и математического мастерства) на примерах: «Хотя во времена, к которым относится наша история, человечество не знало ЭВМ, неизвестный, нашедший это решение, был первым в мире компетентным программистом. Я рассказывал эту историю разным людям. Программистам, как правило, она нравилась, а их начальники обычно сердились все больше и больше по мере ее развития. Настоящие математики, однако, не могли понять, в чем соль.» – это знаменитая история о туалетах⁵⁸.

⁵⁸ <http://hosting.vspu.ac.ru/~chul/dijkstra/pritcha/pritcha.htm>

Помним, что «программирование» – это для Дейкстры практика «структурного программирования», то есть дисциплина/теория алгоритмики на императивном языке с простыми структурами данных. Но вот это «чем мышление программиста отличается от мышления математика» – это оказывается важно, Дейкстра пытался разобраться, чем рассуждения с объектами программистского интереса/внимания отличаются от таковых для математиков и физиков. «Хвост коровы Маргариты – это часть стада» для системного мыслителя неправильное высказывание (нет осмысленных операций в жизни для хвоста в стаде, а вот для «хвоста у коровы»/«хвоста в корове» и для коровы в стаде – есть! Системные уровни важны, через них нельзя прыгать в мышлении!), а для математика, логика, физика – правильное. Системное мастерство по сравнению с математическим, логическим или даже физическим мастерством будут рассуждать по-разному, давать разные ответы на даже простые вопросы! Системный интеллект и математический/логический или даже физический интеллект породят разные варианты какого-то прикладного мастерства, ибо они мыслят по-разному!

Тут произошёл незаметный, но важный сдвиг в онтологическое трансдисциплинарное разбирательство: мы говорим уже не об интеллекте и мышлении, а также мастерстве и рассуждениях как таковых, а об их видах (специализациях), их экземплярах и примерах (классификациях), об их частях (композициях, именно это отношение между объектами-системами на разных системных уровнях), создании и развитии (один объект как-то создаёт и развивает другой объект, часто по цепочке создания). Мы задаёмся вопросом отношений, в которых разные экземпляры и целые множества «интеллекта», «мышления», «мастерства», «рассуждений» могут находиться друг с другом. В онтологии вопрос выбора типа отношения в трудных случаях (например, выбор специализации, классификации или даже композиции) для создания компактной теории/модели/объяснений/онтологического описания зависит от тех проблем, которые вы пытаетесь решить. Для решения каких-то проблем удобно выбрать мир состоящим из одних объектов и отношений между ними, для других проблем – выбрать по-другому. Так что пока не будем обсуждать этот вопрос более подробно, пока вы сами не займётесь исследованиями интеллекта и мышления, мастерства и рассуждения. В любом случае помним, что речь идёт о работающих вычислителях (интеллекте, мастерстве, которые реализуются работающими мозгами, компьютерами и линиями связи) и разворачивающихся во времени в них физических процессах вычисления (мышлении, рассуждениях). Так что интеллект, мастерство выделяются в окружающих людях и их компьютерах и других инструментах вниманием, равно как происходящие в ходе протекания процессов мышления и рассуждений изменения/поведение тоже выделяются из всех изменений в окружающем мире тоже вниманием. А вот куда направлено это внимание, это и определяется трансдисциплинами, занимающимися интеллектом и мастерством, мышлением и рассуждениями.

И, конечно, познание и рассуждение тесно связаны ещё и тем, что в машинном интеллекте обсуждается как «обучение/познание всю жизнь»/lifelong learning: все рассуждения оцениваются на предмет того, насколько они оказались успешными в реальной жизни, и эта успешность или неуспешность тоже идёт как входной материал для мышления. При этом времени на мышление (познание и обучение) не хватает в живой природе, и по итогам рассуждений при действиях во время бодрствования познание идёт ещё и во сне (мозг пересматривает записи того, что там происходило в ходе практики и использованных в практике рассуждений и доучивается: перестраивает мастерство, улучшает его).

Так же рассматриваем мышление и рассуждение в ходе творчества и импровизации (помним, что там обычно участвует какой-то генератор случайностей, меняющий рассуждения), познание с подкреплением, познание на основе принципа свободной энергии (есть и такие объяснения познания живыми существами)⁵⁹.

⁵⁹ https://en.wikipedia.org/wiki/Free_energy_principle

Конечно, мышление в его SoTA варианте (с выходом на осознанность в использовании каких-то новых понятий из новых полученных обучением или исследованиями объяснений/теорий/моделей) в мире встречается сильно реже, чем простые рассуждения. СМД-методологи любят говорить, что «чистое мышление» так же часто встречается в мире, как танцы лошадей. А как же люди занимаются какой-то деятельностью? Они мыследействуют!

Вычислений интеллекта, то есть мышления у человечества по объёму не так много. Это главным образом рассуждения с использованием трансдисциплин (логики, онтологии, системного мышления и т.д.). Но эти вычисления таки бывают. Основной объём «думания», прикладных рассуждений на планете – это мыследействие/вывод/рассуждение по правилам с использованием плодов интеллекта: обеспеченного/enabled интеллектом мастерства как прикладных теорий/дисциплин/моделей/объяснений по решению каких-то классов задач, для которых понятна понятийная структура. Нет затыков в (мысле) деятельности – мозг работает в режиме автомата, лёгкий режим с использованием быстрого интуитивного режима работы мозга-вычислителя S1 (как это было описано в книге Д. Канемана «Думай медленно... решай быстро»⁶⁰). Случился затык, найдена проблема – включается медленный режим работы мозга S2, который за счёт падения скорости и вывода рассуждения в сознание (помним, что сознание управляет вниманием!) гарантирует выполнение правил рассуждения, то есть использование заведомо известных операций с заведомо известными объектами, которые определяются какой-то дисциплиной. Или же такое медленное осознанное рассуждение с использованием трансдисциплин будет в рамках мышления, занимающегося поиском правил для какой-то прикладной дисциплины, которую должен создать интеллект.

У мыслителей, которые главным образом вырабатывают новые понятия (наука, да и существенная часть инженерии) познания/мышления/learning много. А вот у каких-нибудь клерков среднего звена – понятийной работы ноль, сплошные «рассуждения на полном автомате», вот их и списывают за ненадобностью, заменяют компьютерами, это легко. Пока ещё плохо понятно, как заставлять заниматься мышлением компьютер, поэтому интеллект тут берётся у разработчиков софта со всем их искусством исследования рассуждений в ходе выполнения каких-то прикладных практик (методологическая работа) и пересадки найденных правил рассуждений в компьютер (программная инженерия). Но хорошо известно, как потом заставить рассуждать компьютер, когда его уже научили делать рассуждения (то есть «разработали софт»). Софт типа Bing, Bart, прочие «нейросетевые ассистенты» как-то пытаются решать эту проблему полноценного компьютерного мышления, но это ещё не слишком надёжно и плохо работает для ответственных приложений. Из компьютеров пока получаются плохие методологи, они плохо описывают новые деятельности, плохо предлагают новые понятия. В любом случае, ситуация быстро меняется, ибо человеческий и машинный интеллект задействуются не по одиночке, а совместно – и вот эта связка работает уже много надёжней.

⁶⁰ <https://www.ozon.ru/product/dumay-medlenno-reshay-bystro-kaneman-daniel-240690039/>

Бесконечное развитие требует интеллекта

Эволюция заключается в бесконечном развитии, open endedness⁶¹, в выходящем на множество различных масштабов вещества и масштабов времени непрерывном познании⁶². Эволюция глубоко физична, по мере эволюции растёт сложность эволюционирующих систем⁶³ и появляется всё более и более сильный интеллект.

Умность/интеллектуальность появляется в ходе эволюции как раз как средство для ускорения бесконечного развития, для бесконечного прироста видов мастерства агентов (животных, людей, а дальше технических систем и гибридных коллективов из людей и оборудования, включая датацентры с AI), бесконечного прироста в классах проблем, которые научилось решать человечество как коллективный агент. Интеллект невозможные ранее задачи (типа полёт по орбите в космосе вокруг Земли или общение по видеосвязи) превращает во вполне решаемые.

Проекты, где требовались наборы старых навыков и умений большинства людей, старое мастерство, стремительно теряют актуальность – к ним прилетают «сбоку» (из других отраслей) подрывные технологии, и эти проекты заканчиваются. Телеграф вдруг исчезает, и людям с мастерством телеграфиста нужно вписываться в новые проекты, отращивать себе новое современное мастерство – самое разное, часто никак с телеграфом не связанное. **В этот момент никакой интеллект им не будет лишним, ибо сила интеллекта определяет скорость обучения новому мастерству. Если интеллект низкий, то к моменту достижения нужного уровня мастерства нужда в этом виде деятельности может отпасть. Если интеллект у человека высокий, то обучение новой деятельности пройдёт быстро, и останется ещё время это мастерство использовать (а потом всё равно нужда в этом виде деятельности отпадёт).**

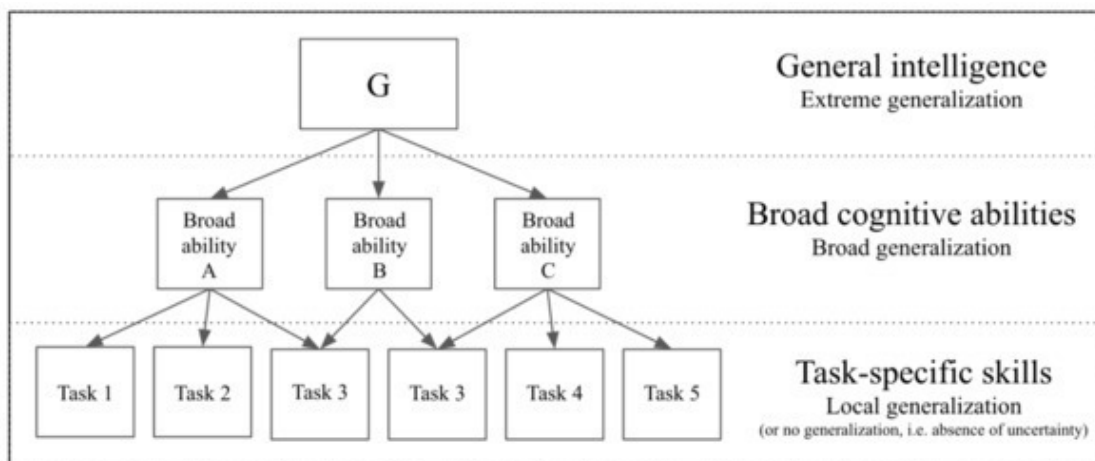
Интеллект тем самым проявляется на задачах, которые не встречались в момент его создания – неизвестны ни самому интеллекту, ни создателю или этого интеллекта (если речь идёт об аппаратуре – мозге людей или программно-аппаратном комплексе AI), ни учителю этого интеллекта (если речь идёт о предобучении аппаратуры – и людей, и AI). Родители не знают, с какими проблемами в ходе бесконечного развития столкнётся их ребёнок, учителя не знают, с какими проблемами столкнётся их ученик, разработчики робота не знают, с какими проблемами столкнётся их робот.

Замерять решение человеком или компьютером (или многими людьми со многими компьютерами) задач какого-то одного узкого класса, чтобы определить силу их интеллекта – неправильно. Нужно измерять способности (broad abilities) к освоению новых предметных областей, то есть скорость приобретения мастерства/skills в решении проблем в этих предметных областях.

⁶¹ <https://www.oreilly.com/ideas/open-endedness-the-last-grand-challenge-youve-never-heard-of>

⁶² <https://www.pnas.org/doi/10.1073/pnas.2120037119>

⁶³ <https://www.pnas.org/doi/10.1073/pnas.1807890115>



Беря за основу вот эту диаграмму, François Chollet предлагает определять следующие уровни интеллекта по линии универсальности проблем/задач, которые он может научиться решать:

- **полное отсутствие интеллекта**: точно заданные образцы задачи. Заполнение точно известной компьютерной формы значениями, которые берутся из точно известных мест. Переноска заготовок от одного определённого станка к другому определённому станку.
- **локальная генерализация** aka *robustness*: обработка точки в более-менее плотно заданном вероятностном распределении задач – *adaptation to known unknowns within a single task or well-defined set of tasks*. Заполнение анкет разной формы (все возможные формы анкет известны заранее). Переноска заготовок между разными станками (между какими – известно заранее). Это подмастерье.
- **широкая генерализация** aka *flexibility*: разработчик/учитель этого не предвидел, решение широкого класса задач – *adaptation to unknown unknowns across a broad category of related tasks*. Заполнение анкет как таковое, самых разных форм и содержания. Переноска заготовок между всевозможными станками, и не только станками, по потребности. Это мастер, он ориентируется по обстоятельствам.
- **экстремальная генерализация** aka *generality*: как у человека – *adaptation to unknown unknowns across an unknown range of tasks and domains*. Умею заполнять анкеты, переносить заготовки. Вдруг потребовалось управлять синхрофазотроном – это не «задача», это уже проблема! Попотел, но смог научиться. Это талантливый человек, «интеллектуал» (у него сильный интеллект, если научился быстро! Или не очень сильный, если научился, но медленно).
- **универсальность**: генерализация на уровне большем, чем человек – *any task that could be practically tackled within our universe*. Во вселенной есть много проблем, которые человеку и в голову не придут, он с ними не столкнётся. Но интеллект уровня выше человеческого сможет научиться решать и эти задачи, сможет выработать нужные для этого знания, умения, навыки, скиллы, мастерство. Это люди со всеми их компьютерами, а потом сверхлюди (мы не знаем, как люди смогут модифицировать себя, когда они решат текущие проблемы биологического старения и смерти, ограничений в биологическом восприятии текущих органов чувств, ограничений в ловкости и силе текущего человеческого тела).

Машинный/искусственный/компьютерный интеллект сегодня в целом решает задачи локальной генерализации/*robustness*, то есть разбирается в узких предметных областях. Это огромный прорыв по сравнению с тупым роботом, выполняющим заданные операции в заданной последовательности только с определёнными предметами, и даже не классами этих предметов.

Chollet (и ещё множество лидеров AI) призывает решать проблемы, появляющиеся при широкой генерализации/flexibility.

Примерно это же имеют в виду люди, когда говорят о каком-то классе человеческого интеллекта: эмоциональный интеллект (интеллект, разбирающийся с самыми разными проблемами, связанными с эмоциями – что вы будете делать, когда вас захватывает эмоция, с которой ранее вы не встречались?), коммуникационный интеллект (интеллект, который может справиться с огромным разнообразием проблем, встречающихся в коммуникации – будь то в переговорах трёх конфликтующих групп, или даже в разговоре с самим собой), математический интеллект (интеллект, который способен справиться со всевозможными математическими проблемами), и так далее. По факту, это не столько «проблемы» (которые никто не знает, как решать), сколько задачи, которые можно успешно решать, если использовать уже известные людям сегодня знания. Ну, и это бытовая речь: мало что изменится, если заменить «интеллект» на «мастерство»: эмоциональное мастерство, коммуникационное мастерство, хотя вот математическое мастерство уже попадает в серую зону: профессиональные математики, конечно, имеют прикладное математическое мастерство (значительная часть выпуска университетских математиков уходит работать в страховые компании и банки, где они занимаются актуарными расчётами⁶⁴), но всё-таки математики формулируют проблемы и находят новые способы их решать, речь всё-таки идёт именно о математическом интеллекте как решателе проблем (интеллект находит способ решения какого-то класса задач, который непонятно как решать – умение/мастерство решения этих задач является результатом его работы). Так что в случае математического интеллекта наше онтологическое чутьё подсказывает, что это всё-таки что-то другое, чем математическое мастерство. В случае кулинарного или эмоционального мастерства наше онтологическое чутьё молчит, мы понимаем, что бытовым языком тут волен использовать какие угодно слова «для красивого словца».

AGI (artificial general intelligence, искусственный универсальный интеллект) – так называют дисциплину инженерной практики создания небиологического вычислителя для мышления. Нынешняя цель AGI – создать интеллект широкой генерализации/flexibility, в котором он потенциально может выработать мастерство в решении тех же проблем, которые мог бы научиться решать биологический человек, а не кошка или какое другое животное. Обратите внимание на формулировку, включающую в себя возможность относительно бесконечного развития: речь идёт не об умении специализированного на каком-то классе задач «искусственного мастерства» решать задачи так же, как обученный этому человек. Эта формулировка про «такое же решение задач» не включает в себя развития. Формулировка про «мог бы научиться решать человек» включает в себя бесконечное развитие, есть ещё множество проблем, которые люди ещё не научились решать, и о которых, возможно, ещё они не знают – но можно ожидать, что они это делать научатся (с помощью компьютеров, или без них).

Насколько это развитие бесконечно? Понятно, что человек сам по себе может научиться решать только конечное число классов задач. Но вместе с AI он может изменить и свою биологическую природу, и техническую природу AI (скажем, сегодня ожидается резкий скачок в скорости вычислений при переходе к универсальным алгоритмам на квантовых компьютерах). Плюс учиться решать задачи может человек не только в одиночку, но и целой группой, а хоть и целым человечеством – наука и производство сегодня глобальны, в них участвуют люди по всей планете плюс огромное количество оборудования/аппаратуры и компьютеров.

Статья о бозоне Хиггса вышла с 5154 авторами⁶⁵, столько людей приняло участие в решении этой задачи. Статьи, в которых расшифровывается геном каких-то организмов, у биологов выходят с числом авторов больше тысячи. Интеллект как свойство научиться что-то делать

⁶⁴ https://ru.wikipedia.org/wiki/Актуарные_расчёты

⁶⁵ <https://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567>

новое/решать новые классы/виды задач существует не только у отдельных людей, но и у каких-то коллективов, в том числе включающих в себя людей и компьютеры, в том числе и у всей цивилизации в целом вместе со всем возможным оборудованием. Да, если брать все вычисления человечества, то можно говорить о совокупном интеллекте человечества! Интернет позволяет легко собрать вычислительные мощности и людей, и компьютеров, а потом после решения проблемы предоставить результаты огромному числу других людей и компьютеров, вновь найденное мастерство быстро распространится по планете.

Цель всей деятельности по усилению интеллекта как людей, так и машин – создать сверхчеловеческий универсальный интеллект за пределами человеческой экстремальной силы/универсальности/генерализации/generalality. Такой интеллект сможет решить те классы задач, которые человечество пока не научилось решать. **Такой интеллект экстремальной силы/универсальности не только сможет помочь людям стать биологически бессмертными, наладить межпланетные и межзвёздные путешествия (это задачи, которые нам могут прийти в голову прямо сейчас), но и в рамках бесконечного развития сможет поставить интересные проблемы, чтобы их решать и тем самым продолжить эволюцию за пределы чисто человеческой мечты.** Особо обратим тут внимание, что универсальный машинный интеллект тут не представляется обязательно антропоморфным/парохиальным/земным, также не предполагается «видовое противостояние» между «биологическим видом человека» и «технологическим видом AGI». Нет, мы считаем, что люди друг с другом, а теперь и с компьютерами живут в симбиозе. Но оставим эти рассуждения философам.

Конечно, как любая сложная система (помним, что интеллект мы рассматриваем как мастерство познания в незнакомой ситуации) интеллект имеет ещё множество других характеристик кроме общности. Из наиболее интересных тут являются характеристики вменяемости/persuadability как мера изменений, нужная для рационального изменения поведения системы⁶⁶. Невменяемые часы придётся переделать, кошку можно надрессировать, а людям (и вот сейчас AI) можно что-то сказать – и они изменят поведение. Дальше по этой линии идёт обсуждение prompt engineering⁶⁷ и даже нейролингвистического программирования/neuro-linguistic programming⁶⁸ (при этом обращение нейролингвистического программирования к «бессознательному» сегодня считают просто учётом характера человеческой нейросети, распознающей какие-то паттерны и реагирующей на эти паттерны).

⁶⁶ <https://www.frontiersin.org/articles/10.3389/fnsys.2022.768201/full>

⁶⁷ https://en.wikipedia.org/wiki/Prompt_engineering

⁶⁸ <https://anlp.org/knowledge-base/definition-of-nlp>

Интеллект определяет скорость обучения новому мастерству

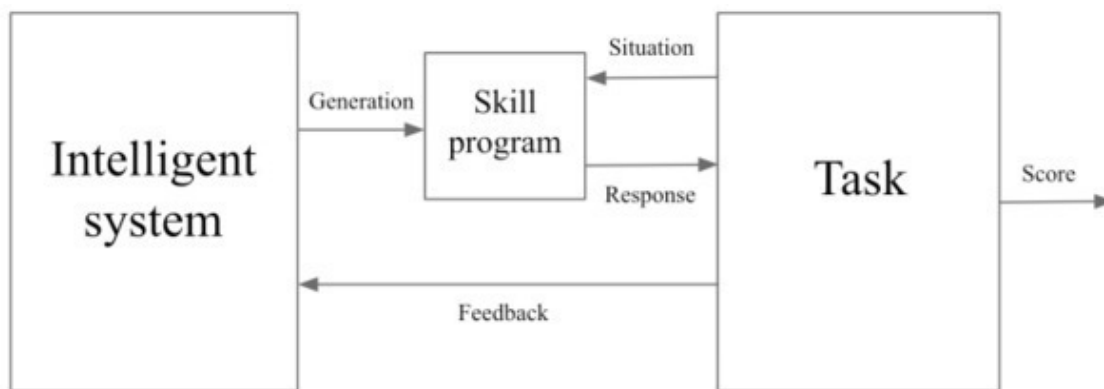
Основные отличия человеческого интеллекта от машинного интеллекта представлялись ещё несколько лет назад ровно в степени его общности/универсальности/generality, поэтому отсылка к интеллекту, который «такой же умный и вменяемый, как человек» обозначалась как artificial general intelligence. Сначала считалось, что general – это примерно «умный как человек-школьник». Поэтом незаметно стало считаться, что это «умный как средний человек», потом – «умный как средний профессор», потом – «умнее человека». После чего оказалось, что технология больших языковых моделей даёт достаточную степень общности в предметных областях (но не в типах решаемых проблем!), чтобы вот это AGI превратилось в просто AI как указание на «машинное происхождение». Заодно оказалось, что AI при помощи технологии больших языковых моделей⁶⁹ не учится действовать в мире как Маугли, взаимодействуя с теми объектами окружающего мира, что случайно встретятся в природе. Нет, познание мира большими языковыми моделями проходит так же, как у людей: их «насиленно встречают» с описаниями самых разных частей мира, имеющих в текстах. Грубо говоря, большие языковые модели учатся так же, как и люди – «в школе, в университете, читая книжки». Люди и AI для получения своего интеллекта «с нуля» знакомятся с огромным объёмом книжного знания, а не просто ощупывают и осматривают окружающий мир. Разница только в том, что AI знакомится с огромным объёмом текста «по всем наукам», а человек знакомится с небольшим объёмом текста по избранным предметам, а потом добирает специализации в конкретной предметной области уже после вуза и школы.

Мы хотим специально организованным предобучением примерно бакалаврского уровня усиливать человеческий интеллект, повышая степень его широкой универсальности/генерализации/flexibility, хотя это на ступеньку меньше, чем «теоретическая» человеческая экстремальная универсальность.

При этом мы не будем забывать о ходе на универсальность через симбиоз человека с компьютерами, то есть ходе на киборгизацию, включение экзокортекса. Скажем, человек обладает биологически плохой памятью и в силу этого сниженным интеллектом – но ведение дневника даже на бумаге и тем более в компьютере поможет помнить много и неограниченно долго. Библиотека с полнотекстовой поисковой системой ещё лучше решает проблему с памятью. Человек медленно умножает десятизначные числа – инструмент-калькулятор ему в этом поможет, а программируемый калькулятор как внешний вычислитель (инструмент!) и подавно. Человек с книгой и калькулятором сможет научиться решать задачи, требующие памяти и вычислений быстрее, чем человек без книги и калькулятора. Человек с книгой и калькулятором тем самым будет умнее человека без книги и калькулятора. А человек с современным даже не компьютером, а дата-центром умнее, чем человек с книгой и калькулятором. А группа людей со множеством дата-центров вообще оказывается умнее всех одиночек с компьютерами. Вы поняли идею: мы не верим в усиление чисто человеческого интеллекта, поэтому предобучать будем сразу людей с их компьютерными экзокортексами. Отдельный вопрос, что тут происходит с вменяемостью: если группе людей дать много разных инструментов (например, баллистических ракет с ядерными боеголовками), то вероятность того, что вы рационально уговорите их изменить своё поведение, неожиданно может снизиться, а не увеличиться.

Chollet даёт вот такую диаграмму, определяющую интеллект:

⁶⁹ https://en.wikipedia.org/wiki/Large_language_model



По этой функциональной диаграмме интеллект/интеллектуальная система создаёт умение что-то делать как отдельное мастерство/умение/прикладное_знание/«программу скилла», и уже это мастерство/умение решает каждую отдельную задачу, потихоньку превращаясь в нетрудный для выполнения навык («автоматизируясь» через большое число повторений, уходя в бессознательное и освобождая ресурс внимания). Интеллект – это вычислитель со способностью выработать мастерство/умение, переходящее постепенно в навык, то есть исполняющееся без сознательного к нему внимания. **Не можешь чему-то научиться за приемлемое время – это тебе не хватает интеллекта, какого-то входящего в состав интеллекта мыслительного мастерства!**

Котёнок может быть очень умным для котёнка, но не способным научиться играть на рояле. Поэтому у котёнка мы считаем интеллект слабым по сравнению с человеком (но сильным по сравнению с рыбой). Если человек оказывается неспособным научиться играть на рояле, неспособным научиться математике, неспособным научиться операционному менеджменту, и так далее по всем видам задач – мы его не будем считать очень умным, откажем ему в интеллекте. Люди-мнемоники в цирке умеют в уме умножать десятизначные цифры, в этом они не хуже калькулятора. Или помнить бессмысленный длинный текст, не хуже книжки. Мы их не считаем особо умными, если они не демонстрируют, что они могут выучиться чему-то ещё. Калькулятор или книжку мы не ценим за их интеллекты.

Если человек постоянно демонстрирует способность освоить какую-то новую предметную область (универсальность! Сила интеллекта в его универсальности: скорости освоения самых разных новых задач!), поднимая и поднимая сложность решаемых им проблем, мы говорим, что у этого человека сильный интеллект. Если человек научился решать один класс задач, но не в состоянии выучиться чему-нибудь ещё, интеллект его будет считаться слабым (неуниверсальным! Малая скорость освоения нового, времени на новое требуется столько, что жизни не хватает!) – независимо от того, насколько сложны те немногие задачи, которым этот человек смог научиться. Этот человек может считаться уникалом, артистом цирка, рекордсменом Гиннеса – но не обладателем сильного интеллекта.

Интеллект связан с универсальностью в части классов решаемых задач и скоростью обучения их решать, а также с вмещаемостью как способностью изменять своё поведение рациональным образом на основе получения информации из текста (речи, книги, выдачи компьютера). Единственный способ подтвердить интеллект – это продемонстрировать, что ты обучаешься решать всё более и более сложные новые проблемы, а также внимаешь рациональным аргументам для изменения своего поведения. Например, научиться арифметике, потом высшей математике, потом инженерным вычислениям, потом вычислениям универсальных алгоритмов, и так далее – до бесконечности усложняя и меняя виды проблем, классы задач и исправляя ошибки, если на них тебе указы-

вают. Если ты просто демонстрируешь решение одного класса задач, вновь и вновь решая арифметические задачи и не двигаясь дальше, не исправляя ошибки и не реагируя на аргументы, то интеллект не будет задействован, он так и будет считаться слабым, «достаточным только для арифметики» и «механическим в своих проявлениях» по линии вменяемости.

#

Интеллект врождённый и приобретённый

Сам Chollet предлагает шкалу универсальности в решении разных классов проблем как силы интеллекта использовать для оценки систем сегодняшнего машинного/искусственного интеллекта. Люди не работают голыми мозгами в разработке чего бы то ни было, они задействуют компьютеры – системы автоматизации проектирования, программы имитационного моделирования, нейронные сети как универсальные аппроксиматоры и т. д. В своей работе по измерению силы интеллекта Chollet выделяет такие подсмотренные у человеческих младенцев элементарные функции как

- умение выделить объект по связности в его представлении в окружающем мире,
- отслеживать этот объект в мире при его перемещениях,
- отслеживать влияние объектов друг на друга,
- умение преследовать какую-то цель,
- умение считать,
- какие-то умения в области геометрии и топологии – типа распознать симметрию в объекте, или выделить прямую линию или прямой угол.

#

Эти **врождённые способности** как частное мыслительное мастерство (а интеллект, как мы помним, состоит из широких/трансдисциплинарных способностей!) и составляют по его мнению «аппаратную» основу человеческого интеллекта, остальному люди учатся с использованием этих врождённых способностей. Другие исследователи соглашались, что какие-то функции у человека как носителя сильного/широкого интеллекта реализованы аппаратно лучше и связывают их со сложной структурой мозга, которая оказывается связана ещё и с генами, кодирующими microRNA⁷⁰. Геном – это тоже «софт», который «исполняется», приводя к разворачиванию полноценного человеческого мозга (или осьминожного мозга, хотя там интеллекта меньше, но больше, чем у мозга тараканов – аппаратура таки важна!). А затем на этой аппаратуре реализуются те или иные «виртуальные аппаратуры», алгоритмы интеллекта. Как любят повторять специалисты по компьютерным архитектурам, «граница между программным и аппаратным обеспечением обычно размыта».

Мы согласны с Chollet, что у выросшего в цивилизованном мире человека интеллект состоит из:

- **врождённых способностей/мыслительного мастерства**, которые «аппаратно» имеются в мозгу человека и определяются генетически, являются результатом биологической эволюции. Эти врождённые способности могут быть использованы как основа для дальнейшего усиления интеллекта через предобучение трансдисциплинарным рассуждениям. Простые тесты из набора IQ должны быть связаны именно с врождёнными способностями, хотя на деле это и не соблюдается. Врождённые способности определяются генетически, и не так много можно сделать, чтобы их усилить обучением, хотя мозг пластичен и в какой-то мере может менять свою структуру для упрощения решения каких-то часто встречающихся задач. Кошку не научишь читать, сколько ни учи, речь об этом. Человека тоже научить можно явно не всему. В любом случае, речь идёт об интеллекте, именно поэтому **про детей с большим IQ говорят «талантливый в одном будет талантлив и в другом», это прямо совпадает с определением сильного интеллекта: «универсальный талант», а не «талант к одному классу задач».** Это и есть тот самый «фактор G», фактор самых общих способно-

⁷⁰ <https://www.science.org/doi/10.1126/sciadv.add9938>

стей к обучению, доступных человеку. Дальше можно обсуждать, насколько это должно сопровождаться какими-то другими наследуемыми способностями. Например, усидчивость оказывается связана с талантом⁷¹: кому-то скучно потратить на какое-то действие 10 часов, а кому-то нет – и вот этот второй при том же интеллекте вдруг получает дополнительное преимущество, его нейронная сетка научится что-то делать лучше при той же аппаратуре, и это тоже наследуемое свойство!

• **Выученных/приобретённых способностей/мыслительного мастерства**, получаемых предобучением каким-то трансдисциплинам. **Приобретённое мыслительное мастерство отличает людей с хорошим образованием от людей с плохим образованием: они оказываются «более талантливыми» (потому как правильно образованы, а не потому образованы, что оказались более талантливы!).** Люди с хорошим образованием могут потом выполнить быструю подстройку своих знаний под новый проект, быстро освоить новое мастерство, разобраться с новым делом. А то и без подстройки: если окажется, что речь идёт об использовании каких-то универсальных умений (трансдисциплин), то и без подстройки можно справиться. А с плохим образованием люди тоже могут разобраться с новым делом, но это происходит медленно, их интеллект слабей. Почему медленно? Потому что им приходится не просто подстраивать свои знания, им приходится ещё для этого и дополнительно предобучаться, часто очень неоптимальным образом, без использования трансдисциплин как накопленного цивилизацией опыта предыдущих поколений. Представьте, что взрослый дикарь приехал из джунглей, где он только охотился и собирал растения. Сколько времени ему нужно потратить, чтобы стать инженером? Он даже в вуз пойти сразу не сможет, ведь у него не будет даже школьных знаний! Речь сразу идёт о многих годах, которые люди тратят на обучение трансдисциплинам. Это ничем не отличается, по большому счёту, от обучения нынешних версий AI, которых сначала долго и много учат «в школе», чтобы получить «большую языковую модель» (large language model, это обучение pretraining), затем обучают их более узким предметным областям (это finetuning), и только затем уже обучают совсем узким условиям ситуации, давая им какое-то задание с подробным описанием (in context training, prompt engineering).

Отдельно нужно обсудить: а можно ли вот так накапливать знания, передавая их от чему-то самостоятельно научившихся людей и AI к ещё не научившимся, да ещё и не лично, а через главным образом разные тексты с редкими картинками (даже не видео)? Можно ли целенаправленно провести «предобучение» для людей, грубо говоря, не заставляя их сразу «жить и работать», а обучая в школе и вузе? Или же каждый человек должен накапливать все знания «на опыте жизни», как-то самостоятельно? Были проделаны эксперименты, показывающие, что передача знания от поколения к поколению вполне возможна, и эта передача идёт на естественном языке, которого оказывается вполне достаточно. Необязательно учиться всему «с полного нуля», набивать себе собственные шишки на собственных неудачах, теряя на это много времени, можно получить опыт современников или даже предыдущих поколений из культуры, в том числе получить нужное знание через текст⁷² – и сразу начинать приобретать новый опыт, которого ещё не имели предыдущие поколения исследователей мира, предыдущие поколения инженеров, менеджеров, предпринимателей. И ровно то же самое происходит с искусственным интеллектом, все современные «умные чат-боты» учатся на огромных наборах текстов прежде всего.

В принципе, огромное число проблем можно решать просто методом перебора разных вариантов решения (оставим вопрос о качестве воображения, чтобы предлагать достаточное число и разнообразие вариантов). Этот метод перебора называется **методом проб и ошибок**.

⁷¹ <https://www.economist.com/science-and-technology/2014/07/05/practice-may-not-make-perfect>

⁷² <https://arxiv.org/abs/2107.13377>

Это основной метод работы многих и многих людей, tinkering/возня как в «он возится с автомобилем», это подчёркивается в книге Нассима Талеба «Антихрупкость». Но возня/«метод проб и ошибок» срабатывает увы, за огромное время и с потреблением огромных материальных ресурсов. Ещё ведь придётся найти то, что нужно будет перебирать, заранее ведь это тоже неизвестно – и перебирать приходится по огромным цепочкам создания. Вы бы догадались, что антибиотики помогают против бактерий в те времена, когда само понятие бактерии было ещё неизвестным? Проблема поиска антибиотиков не могла быть даже поставлена! Догадались бы, что надо использовать радиотриод в качестве логического элемента в вычислительной машине времён Бэббиджа, чтобы получить электронно-вычислительную машину, а не механо-вычислительную или пневмо-вычислительную? Время «возни» можно резко сократить, если возиться с какими-то уже известными из культуры предметами (например, «возиться с микропроцессором», а не возиться с очищенным кремнием в надежде, что в итоге этой возни появится какой-то компьютер, или возиться с разными сортами стали, в надежде, что когда-то из этой возни появятся огромные стальные ракеты Starship и Super Heavy. Нет, «с чем возиться» в методе проб и ошибок тоже зависит от уже накопленного человечеством знания.

Многие сегодняшние проблемы не могут быть решены сегодняшними плохо сконструированными (а эволюция ведёт к отнюдь не оптимальным «врождённым» решениям по части интеллекта⁷³!) и плохо обученными (образование в мире отнюдь не идеально) людьми и машинами. Так что нужно усиливать интеллект, чтобы продолжать эволюцию (как техно-эволюцию, так и биологическую) и исправлять замеченные ошибки.

Представьте, например, что мы ещё не знаем, что такое «свет», а ведь первые микроорганизмы этого не знали! Или не знаем, что такое спин⁷⁴ (который используется в спинтронике⁷⁵), про который догадались только в 1924 году, меньше ста лет назад. Если мы мало знаем о структуре мира, то требуется огромное время интенсивных выходящих в мир для проведения экспериментов рассуждений, чтобы узнать о каких-то проблемах, а затем их решить. И ещё надо узнать о правилах рассуждений, которые ведут к рассуждениям без ошибок, логика у человечества тоже прошла долгий путь развития.

Если мы хотя бы частично что-то знаем о структуре мира (всегда частично, всегда мало, даже через десять тысяч лет это будет «частично» и «мало», развитие бесконечно!), это бы в десятки, тысячи, миллионы раз уменьшило количество вычислений/мышления интеллекта по выработке мастерства в решении связанного с этой особенностью структуры мира класса задач.

Скажем, какую-то проблему мы можем решить человеческим мозгом за десять тысяч лет интенсивных размышлений. Это побольше, чем время существования человеческой цивилизации. Но если мы сделаем какие-то удачные догадки/гипотезы/guesses/предположения о структуре задачи и её предметной области, и они снизят объем вычислений в десять тысяч раз, то проблема будет решена всего за год. И можно будет переходить к следующим, более сложным проблемам.

Ускорение в десять тысяч раз по сравнению с «вознёй» возможно? Бывает ли ускорение на порядки величины по сравнению с «обычной скоростью решения задач»? Да, бывает! Так, квантовые компьютеры уже в определённых классах алгоритмов несравнимо (на много порядков величины) быстрее классических компьютеров, и это квантовое превосходство/quantum

⁷³ Посмотрите примеры неоптимальности эволюционных решений. Всё работает, но крайне неэффективно, ибо наследуются какие-то черты конструкции из предыдущих поколений. Скажем, у рыб не было шеи, и гортанный нерв от мозга к горлу шёл по оптимальной прямой траектории, у человека с этим уже плохо и не оптимально, нерв «возвратный», ибо проходит через петельку кровеносных сосудов около сердца а у жирафа такая петелька вообще вызывает удивление своей неоптимальностью: https://ru.wikipedia.org/wiki/Возвратный_гортанный_нерв. Ни один рациональный конструктор такого бы не допустил!

⁷⁴ [https://en.wikipedia.org/wiki/Spin_\(physics\)](https://en.wikipedia.org/wiki/Spin_(physics))

⁷⁵ <https://en.wikipedia.org/wiki/Spintronics>

supremacy⁷⁶ быстро увеличивается. Или в 2021 году было предложено ускорение на несколько порядков скорости обучения игры в видеоигры для алгоритмов обучения с подкреплением, и были достигнуты скорости обучения примерно такие же, как у человека. Буквально десяток лет назад речь шла о проблеме, которая вообще не решалась, компьютер не мог обучаться игре в видеоигры! Потом мог обучаться, но требовались огромные вычислительные мощности, и дело было хуже, чем у человека примерно в десять тысяч раз, требовался суперкомпьютер. И вот задача решена предложением нового алгоритма, использующего догадки о структуре знаний при игре⁷⁷.

Цивилизация (и особенно в ней наука, она ровно этим и занимается) даёт нам разной степени удачности общие предположения о структуре абстрактного (математические объекты) и физического мира и учит формулировать проблемы. Это приобретённый, выученный интеллект: он позволяет решать задачи в десятки тысяч (а то и более) раз быстрее, чем это могло бы быть сделано необученным структуре окружающего мира интеллект как «аппаратной» частью мозга «дикого» человека, не получившего образования. Цивилизованный человек, мозг, интеллект (это всё вложенные части, в быту мы используем все выражения) – это обученный, образованный человек, мозг, интеллект. Цивилизованный интеллект (мозг, человек) содержит в себе не только врождённые мыслительные способности, врождённое мыслительное мастерство, но и приобретённое/выученное. Интеллект цивилизованного человека оказывается не таким уж естественным: часть его «аппаратна», но часть «программна», прошита цивилизацией в мозгу – это ничем не отличается от любого другого вычислителя. Интеллект смартфона тоже есть врождённый (аппаратный, от микропроцессора конкретной марки), а есть приобретённый – от прошивки производителя, и от конкретного мастерства его прикладных программ. Другое дело, что интеллект смартфона очень слабый, ибо микропроцессор его очень ограниченной производительности, даже с учётом того, что в современных моделях смартфонов используются аппаратные ускорители для нейросетей, да ещё и алгоритмы прошивок абсолютно не универсальны в части возможности решения разных классов проблем.

Помним, что сила интеллекта в его универсальности, а для универсальности нужна скорость работы вычислителя и разнообразие его алгоритмов: есть теорема отсутствия бесплатного обеда/по free lunch theorem, в которой говорится, что один алгоритм не может быть универсально эффективным для всех классов задач, поэтому для универсальности требуется много разных алгоритмов работы вычислителя. Об этом подробнее говорится в книге Педро Домингоса «Верховный алгоритм», которую мы рекомендовали для начального знакомства с подходами к конструированию машинного интеллекта как вычислителя с универсальным (master, верховным) алгоритмом.

Итого: приобретение нового мастерства и у человека, и у AI, и у компании не через чисто «природную смекалку» человека, AI или коллективную смекалку людей и компьютеров в компании, а через «облагороженную образованием смекалку», через получаемые из культуры путём «импорта» готового знания о структуре мира и структуре задач – и уже к этим «импортированным» знаниям предобучения добавляется «возня»/tinkering, «опыт».

⁷⁶ https://en.wikipedia.org/wiki/Quantum_supremacy

⁷⁷ <https://arxiv.org/abs/2107.12544>

Трансдисциплинарный интеллект-стек

Мастерство/умение и навык/скилл/skill – это вычислители для рассуждений по какой-то прикладной дисциплине или трансдисциплине, интеллект – это набор таких вычислителей по разным видам мыслительных практик, поддерживающих рассуждения с объектами и по правилам/объяснениям трансдисциплин этих практик, и с использованием необходимых для этого инструментов. Инструменты тут чаще всего – моделеры, использующиеся для «усиления памяти», даже ручка-бумажка, но иногда для усиления именно вычислений – компьютерные имитационные модели или даже просто калькуляторы.

Трансдисциплины – это и есть сведения о структуре мира, которая оказывается удобной для практик скоростного мышления, мыслительного мастерства быстрого разбирательства с новыми ситуациями. Трансдисциплины – это дисциплины о дисциплинах, наиболее общие мыслительные шаблоны о более конкретных мыслительных шаблонах, используемых для каких-то более конкретных предметных областей. Логика позволяет обсуждать, логичны ли рассуждения какой-нибудь астрологии или квантовой теории поля, онтология позволяет обсуждать объекты мышления в машиностроении и менеджменте, и так со всеми трансдисциплинами.

Проблема, которая займёт всё время очень смекалистого дикаря на полжизни, у обученного мышлению с использованием трансдисциплин человека может занять несколько часов, или даже несколько секунд (особенно, если такой человек будет использовать компьютерный экзокортекс).

Трансдисциплин множество самых разных, они выстроены в условный стек («стопку»), поскольку внутри рассуждений о каких-то одних типах объектов одних трансдисциплин будут использованы рассуждения о других типах объектов других трансдисциплин. Мы называем такой условный (потому как там всё-таки полноценный граф, а не какая-то последовательность, но в целях упрощения мы это игнорируем и продолжаем говорить о «стопке») стек трансдисциплин, использующихся для рассуждений о самых разных предметных областях, в том числе предметных областях друг друга, **интеллект-стек**. Приведём его в обратном порядке, снизу-вверх, чтобы было понятней, как одни трансдисциплины пользуются в своих объяснениях уже введёнными другими трансдисциплинами объектами:

- **Понятизация** учит выделять какие-то типизированные (тут явное забегание вперёд: понятие типа будет определено в интеллект-стеке позже, но мы предупреждали об условности предлагаемой последовательности практик) фигуры из фона и делать их предметами рассмотрения, давать какие-то имена этим фигурам. Роль – поэт.

- **Собранность** учит удерживать во внимании «объекты», которые уже вытащены понятизацией. Это делается не «чистым мозгом», а при помощи внешней аппаратуры памяти и поиска в ней. Так что роль – «собранный», и этот собранный – киборг. Впрочем, интеллект-стек относится и к AI, так что «киборг» тут условно, только для людей, чьё внимание усилено компьютерными средствами.

- **Семантика** учит связывать физические/реальные объекты с математическими/абстрактными/ментальными/идеальными, а также работать со знаками, обозначающими объекты. Если вы вытащили своим вниманием объекты из пестроты окружающего мира, можете удержать их во внимании, то дальше можно обсуждать эти объекты, представляя объекты знаками. Роль – семантик.

- **Математика** учит тому, какие бывают «ментальные» объекты, как они могут себя вести, каким образом конструируются одни из других. Роль – математик.

- Физика учит тому, какие бывают физические объекты в реальном мире, а также каким образом мы используем математические/ментальные объекты с хорошо изученным поведением для представления физических объектов с целью рассуждений о них. Роль – физик.

- Теория понятий учит тому, как мы думаем о понятиях – математических/абстрактных/ментальных объектах, которые представляют физические объекты. Человеческий мозг (а значит, и AI, если его научить) представляет понятия или в виде объектов и отношений (теоретическая теория понятий), или как какие-то объекты-прототипы и объекты с описанием некоторых отличий от прототипа (теория прототипов), и это даже не единственные два варианта, есть и ещё. Логика хорошо будет работать с теоретической теорией понятий, а вот метафоры и всякая художественность – с теорией прототипов. Это пригодится для всех последующих обсуждений. Теория понятий учит машинке типов: что все объекты в каком-то смысле подобны друг другу, и это описывается типами. Примеры часто встречающихся типов отношений в теоретической теории понятий – это классификация, специализация, композиция. Появляются и конструктивные теории понятий, где объекты «конструируются» путём каких-то операций, а не просто представляются объектами и отношениями. Роль – типолог.

- Онтология учит отвечать на вопрос, каким способом мы описываем/моделируем мир: как мы определяем важное и неважное (моделирование), как мы используем модели для ответа на вопросы (рассуждения на основе моделей). Мы разбираемся с мета-моделированием (описания как абстракции получаются не произвольно, но абстрагирование управляется абстракцией более высокого уровня), разбираемся с тем, что вещи/системы на разных системных уровнях (то есть уровнях по отношению часть-целое) описываются по-разному, ибо при взаимодействии частей получаются новые свойства (эмерджентность). Модели задействуют понятия (используем теорию понятий, в том числе пользуемся идеей конструктивной онтологии, понятия которой получаются путём применения операций, а не обсуждая отношения «вечных понятий»). А ещё модели используются для проведения по ним рассуждений с целью предсказания будущего состояния мира (демо моделирование/рендеринг/порождение), тем самым после разбирательства с онтологией мы уже готовы заняться рассуждениями и объяснениями. Роль – онтолог.

- Алгоритмика говорит нам о том, как эффективно (с наименьшей затратой ресурсов) вычислять, то есть проводить каким-то физическим устройством (универсальным компьютером) заданные последовательности операций (алгоритмы) над содержимым какой-то памяти, представляющей собой знаки для математических объектов. Алгоритмика тесно связана с понятием интеллекта, так как интеллект – это программно-аппаратно реализованный универсальный алгоритм, способный с большой эффективностью вычислять самые разные функции. Но физическая природа компьютера не позволяет одинаково эффективно вычислять что угодно, а алгоритмика изучает, что же возможно в этом плане сделать на текущей аппаратной базе, какие последовательности операций на той или иной аппаратуре наиболее универсальны и эффективны. Математик, физик, компьютер – это универсальные вычислители, то есть физические объекты, поведение которых как-то отражает поведение математических/идеальных/ментальных/абстрактных объектов. Алгоритмика тем самым и про живых людей с их рассуждениями/вычислениями, и про классические компьютеры с их рассуждениями, и про квантовые компьютеры с их рассуждениями/вычислениями – всё это просто разные типы физики вычислителей. Роль – алгоритмист.

- Логика говорит, какие есть способы вычислений как рассуждений над моделями, дающие наиболее безошибочные результаты: логический вывод, функциональная оценка, вычисления математических функций, интуитивные оценки в человеческом мозге, прикидки, предсказания, и т. д. Онтология для этого уже рассказала про то, как мы нарезали мир на типизированные (или сконструированные) объекты, описав эту нарезку какими-то моделями, алго-

ритмика уже рассказала, что такое рассуждения-как-вычисления, так что методы рассуждений работают как алгоритмы с моделями.

- Рациональность занимается созданием правильных объяснений. Объяснения представляют собой теории/модели, которые рассказывают о причинах и следствиях в физическом мире. К этому моменту, если мы изучали интеллект-стек в последовательности «снизу вверх», из онтологии уже известно про разнообразие моделей, из логики – о разнообразии правил рассуждений. Математика даст возможность оценить формальность работы с причинами и следствиями, физика позволит говорить о соотношении того, что мы представляем рациональными моделями и того, что происходит с реальным миром. Роль – разум, который не приемлет криковых объяснений и нещадно их критикует, а модели использует для принятия решений о деятельности.

- Исследования как практика познания говорит о том, каким образом мы получаем хорошие объяснения. Мы уже понимаем, какие должны быть свойства у хорошего (рационального, на основе моделей) объяснения, и надо теперь объяснить, каким образом мы накапливаем на планете всё более и более точные и удобные в использовании знания о том, как устроен мир. Мы делаем догадки о хорошей объяснительной (причинной) модели/теории, а затем критикуем эти догадки на предмет непротиворечивых результатов рассуждений по этой модели (логика) и на предмет лучшего соответствия предсказаний этой модели с результатами эксперимента (измерения). И дальше та догадка, которая пережила критику (поэтому рационализм у нас – критический!), принимается всерьёз, то есть может быть положена в основу планирования действий. Роль – исследователь.

- Эстетика даёт критерии красоты (в исследованиях принято говорить об элегантности) в результатах мышления и прикладного труда. Эстетика рассказывает, какой отклик вызывает наше поведение не столько в окружающем мире, сколько в самих агентах (и не факт, что современная эстетика обсуждает, например, эмоциональное воздействие каких-то продуктов труда и описаний только на агентов-людей. Нет, современная эстетика рассматривает и агентов с искусственным интеллектом, и искусственную жизнь). Эстетика оказывается такой же трансдисциплиной, как и любая другая. Математики о красивых доказательствах говорят, что они элегантны, «ничего лишнего, всё по делу – это красиво». Менеджеры говорят о lean/элегантном производстве. Это тоже про красоту, один из предметов, изучаемых эстетикой. Красота тесно связана с понятием стиля, «что стильно – то красиво, в стильности проявляется вкус», а понятие стиля – это прежде всего про многоуровневое паттернирование, понятие «похожести произвольного вида». Роль – эстет.

- Этика говорит нам о том, чего нужно добиваться в жизни: какие цели приемлемо ставить агенту и какими средствами добиваться реализации этих целей. Должны ли люди умирать («программируемая смерть» как полезная для эволюции), или лучше бы их сделать бессмертными? А искусственные интеллекты? К этому моменту уже владеем пониманием, что такое объяснения и как устроены исследования – можем теперь разбираться, что делать с результатами всех этих исследований, на что их можно направлять, а на что направлять вроде не следует. Этика тем самым оказывается близкой к понятию стиля, обсуждаемому в эстетике, обсуждает «стиль жизни». Роль – совесть.

- Риторика говорит о том, как убедить какого-то человека совершить какие-то действия. Начинаем с того, что у вас есть какая-то модель ситуации (полученную вами в ходе исследований) и вы имеете перед собой агента, которому вы объясняете вашу модель ситуации и пытаетесь его уговорить использовать эту модель для достижения каких-то совместных его и ваших целей, достижение этих целей вам обоим будет полезно, но агент этого ещё не знает, вы ему объясняете. Но этика вам уже известна, вы не подбиваете агентов (людей, AI, организации, сообщества) на что-то плохое. Риторика существенно связана с тем, что и люди, и AI сегодня представлены нейронными сетями, а не представляют из себя логические машины, поэтому

для них задействуется prompt engineering. Риторика использует свойство вменяемости сильного интеллекта. Роль – ритор.

- Методология рассказывает о труде/деятельности, в которой агенты (раньше – только люди, а сейчас мы подходим к этому не столь антропоцентрично) организовываются в команду, занимают в ней какие-то роли, выполняют работы по каким-то практикам и тем самым добиваются своих целей. Для этого люди и их компьютеры договариваются о работе в какой-то предметной области, логично рассуждают на базе рациональных теорий, ведут коллективные исследования и не теряют внимания к объектам своей работы. Основной объект методологии – «метод работы»/«способ работы»/труд/деятельность/практика, которую выполняют какие-то роли. Методология – наука, «логия», исследует «какие бывают методы работы». Роль – методолог.

- Системная инженерия – это уже нормативная трансдисциплина. Она не столько исследует, сколько на основании методологии и наблюдения за пробами и ошибками в реальном труде предписывает на высоком уровне абстракции (мета-мета-модель из онтологии) устройство деятельности по созданию систем: какие там должны быть практики и какими ролями они потом будут выполняться. Системная инженерия называется системной, ибо исходит из того, что создаются какие-то системы как состоящие из частей-подсистем целые объекты-системы, отделённые от среды, частью которой эти объекты-системы являются. Такое рассмотрение систем изучается трансдисциплинами интеллект-стека, начиная с физики, и далее проходит красной нитью через весь интеллект стек как «системный подход» (рассмотрение мира как набора взаимодействующих систем). Инженерия – это про то, как одни системы (системы-создатели) должны создавать другие системы. Именно «должны создавать» (задаётся норма, «как надо»), ибо «как могли бы создавать» – это методология. Самые разные варианты прикладных дисциплин дальше будут просто специализацией системной инженерии для разных классов систем: если организация, то это будет менеджмент (инженерия организации), если мастерство, то это будет обучение (инженерия мастерства), и так далее. Для рассуждений об инженерных практиках задействуются все предыдущие уровни интеллект-стека (особенно если учесть, что в качестве агентов действуют люди, люди и компьютеры, и даже уже иногда сами компьютеры). Роль – инженер.

Каждая мыслительная практика, основанная на фундаментальной дисциплине/трансдисциплине, при помощи своих понятий и приёмов мышления на их основе помогает разобраться со следующей мыслительной практикой интеллект-стека, а в конце их цепочки – с прикладной практикой, основанной на прикладной дисциплине, которая поддержана какими-то инструментами моделирования (моделерами, специальным софтом для удобного описания ситуаций) или даже просто инструментами (станками, оборудованием или даже какими-то сооружениями, типа нефтяной скважины для практики нефтедобычи).

Без владения трансдисциплинами интеллект-стека трудно понимать тексты, сосредоточиться, рассуждать без грубых ошибок

Без какого-то мастерства собранности вы не смогли бы дочитать до того места в тексте, которое вы сейчас читаете: просто много раз бы отвлеклись и давно забыли бы, чем занимаетесь. Поздравляем, какая-то собранность у вас есть!

Проверим, есть ли какой-то навык в практике теории понятий, навык в исполнении роли типолога. Есть ли у вас какой-то «дребезг» (кинестетические ощущения) в теле, когда вы читаете вот этот отрывок текста?

«Интеллект – это степень универсальности мышления. Интеллект является свойством мышления, отвечающим за новизну, за решение проблем. Как вычислитель, интеллект характеризуется его скоростью».

Люди, которые разобрались в теории понятий и опираются в мышлении на теоретическую теорию понятий (theory theory), будут чувствовать дискомфорт во всём теле, когда читают такое. Теория понятий помогает описать и удерживать при необходимости в сознании «машинку типов», которая разбирается с типами объектов. Давайте посмотрим, что сказано с точки зрения типов в вышеприведённом кусочке:

- интеллект – это степень свойства универсальности мышления
- интеллект – это свойство мышления. Но только что же было, что это степень свойства мышления, а не собственно свойство!
- интеллект имеет своим свойством скорость. То есть если интеллект – это степень свойства, то степень свойства имеет скорость (как это?). Если интеллект – это свойство, то свойство имеет скорость (как это?).

Если «машинка» типов у вас в голове не работает, то эта ахинея будет воспринята вами как вполне нормальный текст, никаких противоречий или странностей вы не заметите, нормальный ведь текст про интеллект, мышление, универсальность, проблемы и даже правда про интеллект как вычислитель! Дальше вы будете рассуждать, имея вот такие туманные представления об интеллекте: «это не-пойми-что, используемое в разговорах о мышлении и как-то относящееся к скорости и новизне, неважно как». Вы будете удивляться, почему ваши вроде как умные рассуждения будут восприниматься другими людьми с работающей «машинкой типов» как ахинея. А ахинею других людей вы будете принимать за чистую монету! У нас на входе курса «Моделирование и собранность» в Школе системного менеджмента были люди, которые честно пытались найти какой-то смысл в бреде, который породила плохо обученная нейронная сетка прошлых поколений. Это не было распознано как бред, то есть порождение больной (в данном случае кремниевой) фантазии!

Мыслительные практики дают возможность понимания чужих мыслей, своих мыслей, нахождения в них ошибок. Трансдисциплины этих практик дают для рассуждений объекты, которые важны, каждая дисциплина – это чеклист таких объектов и чеклист для операций (как рассуждений, так и действий в физическом мире) с ними. Технологии для поддержки этих трансдисциплин дают возможность удерживать это всё во внимании и проверять рассуждения на их безошибочность более надёжно. Если вы не владеете мыслительными практиками, то вы рассеяны (не используете технологии для мышления) и глупы (не используете дисциплины для мышления).

Если вы разобрались с операционным менеджментом как одной из подпрактик системного менеджмента (прежде всего – вы понимаете, какие отношения операционного менеджмента и других практик менеджмента, например, практикой орг-архитектурны), то дальше вы можете за неделю разобраться с конкретными ситуациями, в которых без этих знаний нужно было бы разбираться годами.

Так, если вы пришли на предприятие и обнаружили там специалистов, которые говорят на таинственном языке сил, текущих по жилам, ибо они в какой-то момент впечатлились книжками Алексея Андреева по русской школе менеджмента, то вы сразу распознаете, что речь идёт о «потоке», то есть это операционный менеджмент. Вы очень быстро разберётесь, как говорить на этом языке сил и жил, и как взаимодействовать с остальными людьми в этой фирме. Если у вас в голове уже прошиты общие знания по системной инженерии, а на основе этих знаний вы ознакомились с практиками системного менеджмента как системной инженерии организаций, то вы будете быстро разбираться с самыми экзотическими вариантами прикладных менеджерских дисциплин.

Для инженеров-программистов и «железных» инженеров всё ровно так же: трансдисциплины обеспечивают быстрое разбирательство с прикладными дисциплинами, которые существенно меняются на каждом месте работы. Если ты понял, как создавать всё что угодно (системная инженерия), то как создавать организацию (системный менеджмент) ты поймёшь быстрее, как создать администрацию в организации – с этим тоже быстро разберёшься, как создать софт для администрации в организации – и с этим быстро разберёшься. Будет сразу понятно, что в практике администрирования (финансы, юридическая поддержка, офис-менеджмент, поддержка компьютерных приложений и т.д.) важно, что неважно, чем отличается от других практик, зачем применяются те или иные приёмы работы, там обсуждаемые. Всё понятно сразу, беглость в применении практики наберёте быстро. Если вы не знакомы с системной инженерией в целом, и системным менеджментом как её вариантом в отношении к предприятиям, то быстро разобраться с тем, как создать администрацию предприятия будет невозможно: вроде как будете работать строго по инструкции отдельных практик администрирования, но результаты будут каждый раз не слишком полезные: проблемы будут на стыках с остальными людьми в команде проекта и внешними по отношению к проекту людьми. Скажем, бухгалтерия и юристы будут работать не на благо фирмы, а как враги, останавливая работу, а не помогая работе! Ибо лучшая для бухгалтеров и юристов фирма – это которая ничего не делает, у неё нет налогов, нет затрат, нет рисков! И лучше бы о таких эффектах узнавать из учебных курсов, сразу, нежели «из многих лет опыта ужасной жизни внутри ужасно устроенной компании».

Разные виды мастерства выполнения трансдисциплинарных практик, входящих в интеллект-стек, управляют обнаружением и удержанием объектов внимания в прикладных дисциплинах, определяют рациональные приёмы мыслительной работы с объектами этих дисциплин. Вы легко находите «приглашения» для мышления в непонятных ситуациях, часто даже без подстройки/дообучения. Сначала надо задействовать общие способности к познанию, задействовать мышление сильного интеллекта, чтобы быстро разобраться в новой ситуации, познать/learn происходящее как можно быстрее, а потом уже бегло рассуждать в пределах более-менее понятного прикладного мастерства решения задач в изученной хоть как-то прикладной предметной области. Если ты хороший инженер, и тебе надо вдруг разобраться с менеджментом, то это для тебя будет всего лишь ещё одна инженерная практика. Если ты хороший инженер, и тебе надо вдруг разобраться с инженерией личности (заняться развитием личности), то и это будет всего лишь ещё одна инженерная практика. Хорошо ли это? Конечно, это хорошо! Если четыре уравнения Максвелла компактно описывают весь бесконечно разнообразный мир электромагнитных явлений, то непонятно, почему мы не можем компактно размышлять о бесконечно разнообразном мире создания и развития самых разных систем.

Именно трансдисциплины интеллект-стека работают как алгоритмы интеллекта, когда мы учимся чему-то прикладному, подстраиваем свой мозг к работе в какой-то узкой предметной области в новом классе проектов. При обучении важна сила/универсальность нашего интеллекта, важна вменяемость.

При этом даже чтение и письмо оказываются навыком, которому надо учиться: это навык использования инструмента внешней памяти, усиливающего интеллектуальные возможности «голого биологического мозга». В 21 веке чтение и письмо сами по себе не целевые прикладные умения, а просто предобученные общие способности, нужные главным образом для научения делать что-то ещё. Это в древнем Египте писец владел письменностью как прикладным мастерством. А сейчас это просто результат предобучения, часть интеллекта, это нельзя считать прикладным мастерством/умением/скиллом, просто за умение читать-писать денег сегодня не платят. Если ты похвастаешься «я умею читать и писать», на тебя очень странно посмотрят. Более того, чтение и письмо мы не включаем прямо в состав интеллект-стека (хотя косвенно эти умения разбираются во множестве дисциплин, например, в семантике).

Общим для мыслительного мастерства является мастерство собранности, дающее способность заниматься чем-то одним не просто несколько часов без отвлечения (например, пару часов читать сложную книжку, не отрываясь, без потери внимания – это было обычной нормой образованного человека буквально пару десятков лет назад, хотя и сегодня на планете достаточно людей, которые не утратили эту способность к чтению длинного текста без отвлечения на какие-то чаты и социальные сети), но и несколько лет подряд – удерживая внимание записями, а не «ментально». Более того, есть приёмы удержания внимания целой команды людей, собранность оказывается не чисто человеческой психопрактикой! Мастерство собранности тоже познаётся, а не врождённое. Как и за все трансдисциплины, за мастерство собранности как прикладное умение не платят, эту собранность нужно применить ещё к чему-то. Например, к решению прикладной менеджерской задачи, вроде составления плана-графика конкретного проекта методом критической цепи. Отдельно за задействование мастерства логики не платят! И не платят отдельно за методологическое мастерство! **Не платят отдельно за мастерство по любой из практик интеллект-стека. Платят за прикладное мастерство, но вы его не получите быстро и качественно, пока не овладеете «неоплачиваемым» мыслительным мастерством, пока не станете умным, то есть не получите фундаментальное образование.**

#

Граница между мыслительным и прикладным мастерством размыта

Эволюция делает очень размытой саму границу между врождённым интеллектом в его биологических границах (человеческий детёныш рождается совсем не умным! Но потом растёт и быстро развивается. В том числе окружение этого детёныша и свойство пластичности мозга влияют на то, как развивается биологическая «аппаратная» часть интеллекта) и привнесённым культурой: усиления интеллекта в силу трансдисциплин и прикладных дисциплин. Щенок не умеет в момент рождения ходить, но научается потихоньку и ходить, и бегать, и прыгать – щенки играют с окружающим миром, это их обучение. Люди делают то же самое. Но кроме бегать-прыгать они ещё выучиваются читать-писать, а современный человек по идее должен выучивать ещё в школе, а затем в вузе трансдисциплины, увеличивать силу своего интеллекта, расширять диапазон проблем, которые он сможет решить при встрече с ними. Так что никакой ребёнок не рождается «умным» по абсолютной шкале силы его интеллекта, разве что есть биологическая возможность этим умным стать.

По большому счёту, у людей сейчас не естественный интеллект (вспомним людей-маугли, которые вырастают в джунглях и даже разговаривать не умеют), а вполне искусственный интеллект, полученный обучением знаниям, накопленным поколениями людей – и большинство этих знаний передаются самым умным людям вообще через бумагу и экраны с текстом, а не каким-то другим путём, например, получением «опыта» в решении проблем.

Это и есть «цивилизованность»: **становиться умным от научения, а не просто быть потенциально умным от рождения. Интеллект можно и нужно развивать, а различие между естественным и искусственным интеллектом в этом плане будет несущественным.**

Сам интеллект оказывается не абсолютным, а относительным: он зависит ещё и от прикладных задач. **Интеллект белки, определяемый батареей тестов по разным задачам скакания по деревьям, будет выше, чем интеллект скачущего по деревьям человека. Это нормально.** Другое дело, что их потом могут сравнить по общему числу возможных решаемых проблем, куда попадут и другие даже телесные задачи – и белка тут же проиграет, белка своим телом не сможет делать многое из того, что сумеет сделать человек. Ровно поэтому тесты интеллекта всегда задействуют максимально широкий набор задач, говорят о **батарее тестов** – проверяется универсальность, а не умение решить конкретную задачу. Тест только математики ничего не скажет про ваш интеллект. А вот тест собранности ума и тела, онтологии, риторики, системного мышления, информатики, трудового кругозора – вот такая батарея тестов будет говорить о вашем интеллекте много больше. А ещё могут быть тесты на вменяемость, на агентность. Всё по поводу тестирования интеллекта становится запутанным, на настоящее время не предложено никаких способов тестирования интеллекта: само понятие интеллекта предполагает возможность решения задач, которые не предусмотрены при обучении этого интеллекта. А если задачи предусмотрены, то это тестирование прикладного мастерства.

Интеллект зависит не только от «аппаратуры» устройства AI или человека, но и от учебного плана/curriculum, т.е. порядка предъявления задач, а также от количества информации, которая была предложена в ходе научения, а ещё от качества этой информации. В какой-то момент разработчикам AI казалось, что можно получать всё более сильный интеллект, просто увеличивая размер нейронной сети («объём мозга»), но оказалось, что есть зависимость между размером нейронной сети и количеством данных, которые требуется предъявить этой

нейронной сети, чтобы полностью задействовать размер этой нейронной сети⁷⁸. Поэтому стратегия поменялась: стали применять более простые «аппаратные решения» (алгоритмы нейросетей меньшего размера, требующие меньше вычислительной аппаратуры для работы с ними), но предобучать их на большем количестве данных, а ещё делать лучше сами данные (по двум параметрам прежде всего: уменьшать количество ошибок в данных, чтобы в обучение арифметике не попадали примеры типа $2*2=5$, а также закрывать данными более широкие классы предметных областей, ибо оказалось лучше иметь по паре примеров из десяти предметных областей, чем двадцать примеров из одной предметной области). Оказывается, «образование» кремниевых нейросетей более важно, чем «врождённый интеллект» этих нейросетей. Для людей выполняется всё то же самое. Если потенциального Эйнштейна не учить читать и писать, то он так и останется дикарём, ему не поможет никакая «врождённая гениальность».

Если вам свезло с последовательностью предъявляемых задач, то ваш интеллект может оказаться по их итогам лучше. «Ему повезло попасть в хороший проект, он там сильно вырос за последнюю пару лет» так же важно, как «ему повезло с родителями, они ему дали хорошие гены для мощного интеллекта».

#

⁷⁸ Ключевая работа тут была – <https://arxiv.org/abs/2203.15556>

Важен порядок решения проблем: их надо решать по мере возрастания сложности

Важен не только порядок обучения, но и порядок решения проблем! Проведено много экспериментов, показывающих, что решение какой-то одной трудной проблемы «с нуля» обучающимся вычислителем на базе нейронной сети и какого-то эволюционного алгоритма часто невозможно, но решение какой-то последовательности проблем возрастающей трудности возможно⁷⁹.

Рабочая жизнь – это бесконечное развитие, бесконечное познание, бесконечное обучение. Опыт предыдущих проектов ведь тоже учит мастерству, учит не только образованию, понимаемое как «импорт» уже полученного кем-то знания. Получается, что чем лучше учебный план по обучению мыслительному мастерству/интеллекту – тем сильнее будет получаемый по итогам обучения интеллект! Те, кто решает больше самых разных проблем и берёт их каждый раз на границе тумана будущего, то есть на пределе своих возможностей – у тех вырабатывается не только хороший набор прикладного мастерства, но и сильный интеллект, набор общих для самых разных прикладных умений мыслительных умений/способностей, дающий возможность на следующем такте решить проблему, которую нельзя решить «с нуля», просто потратив много времени именно на её решение. Время на «упереться и попробовать решить» таким не предобученным и неопытным агентом будет потрачено, но результата не будет! Именно это неизменно подтверждается в экспериментах по AI: с нуля агент не может научиться решать какой-то класс проблем, потратив много времени, но если агент накапливает опыт решения проблем промежуточной сложности, то он справляется и с той проблемой, которую не мог решить при попытке это сделать «в лоб», методом грубой вычислительной силы.

Это ровно то, чем отличаются «вечно подающие надежды, но никак не становящиеся мастерами» от в юности середнячков, которые воспитывают из себя настоящих интеллектуалов – и оказываются потом мастерами в прикладных практиках. **IQ⁸⁰, показывающий развитие «аппаратуры» мозга на специфических тестах зрительного восприятия, простых лингвистических тестах, оказывается коррелирующим с уровнем интеллекта, но не определяющим его! Мозг пластичен, синаптические связи в нём отращаются, кровоснабжение адаптируется, интеллект вполне тренируем – как мышцы. Предобучение как хорошее образование перед занятием каким-то делом, правильно подобранная последовательность задач для этого предобучения, и правильно подобранные для этого обучения трансдисциплины влияют на интеллект даже больше, чем родительские гены.**

Вы должны усиливать свой интеллект обучением себя по двум направлениям:

- **обучаясь трансдисциплинам и давая себе деятельностный кругозор, то есть делая «прошивку интеллекта» в её самой современной версии, доступной человеческой цивилизации.** Ещё пару веков тому назад такие трансдисциплины включали в себя чтение-письмо-арифметику. Нынешнее поколение чтение-письмо-арифметику не считает даже за отдельные дисциплины (но таки в начальной школе их учат!), а разные исследователи говорят о разных наборах этих дисциплин, мы в нашем курсе предлагаем свою версию.

⁷⁹ we show in the technical paper that with just 30 minutes of focused training on a newly presented complex task, the [pretrained on multiple tasks] agents can quickly adapt, whereas agents trained with RL from scratch cannot learn these tasks at all, <https://deepmind.com/blog/article/generally-capable-agents-emerge-from-open-ended-play>

⁸⁰ https://en.wikipedia.org/wiki/Intelligence_quotient

• **обучаясь полагаться в том числе и на внешние компьютерные средства просмотра за личным и коллективным вниманием на личной и коллективной внешней памяти, т.е. улучшая свою «аппаратуру» до самой современной версии, доступной цивилизации. Да, по факту нужно становиться киборгом. Вы уже не расстаетесь со смартфоном, признайтесь себе в этом! А смартфон подключает к вам огромные датацентры, которые начинают работать на вас, усиливая вашу аппаратуру мозга.** Прошлое поколение усиливающей интеллект аппаратуры включало в себя ручку-бумажку, которые резко увеличивали объём памяти и внимания, нынешнее поколение такой аппаратуры – дата-центры (основные вычислительные мощности планеты в них!), настольные компьютеры, планшеты, смартфоны и средства связи, которые на предыдущем такте технологической эволюции люди и представить себе не могли. **Человек с луком и стрелами может добыть не дающегося голым рукам зверя, человек с доступом к дата-центру может решить не дающуюся голым мозгам задачу.**

Повторим: уровень IQ⁸¹ не так уж и важен для жизненного успеха. «Коэффициент интеллектуальности» – это просто умение решать какие-то классы задач, замеряемое в простых тестах. Конечно, на IQ влияет в том числе и генетика, но и обучение тоже (мозг пластичен! Если его упражнять, то он развивается). IQ дикарей меньше, ибо их мозг меньше тренировали в детстве. Более того, поскольку в 2020 году во всём мире принимались активные «антиковидные» меры по ограничению общения людей, их буквально запирали дома, как в тюрьме, возможности малышей получать новую информацию были существенно снижены. И для младенцев средний IQ (вербальные, моторные, когнитивные навыки) в США упал со 100 до 78⁸² именно из-за недостатка общения, стимуляции. Это непонятно, навсегда или временно, например, умение говорить должно быть освоено детьми, в более взрослом возрасте оно уже не осваивается, если человека выкормили животные и некому было его учить разговаривать. Будем надеяться, что это отставание будет навёрстано.

Но в любом случае, IQ оказывается не сильно связан с успехом в жизни: упорность и настойчивость в сочетании с удачливостью оказываются более важны. А с учётом того, что люди с низким IQ могут использовать в своих размышлениях результаты умственного труда других людей (в том числе с высоким IQ), и даже результаты вычислений компьютера, то разница и вообще становится небольшой.

Человек с низким IQ, которого выучили умножать в столбик и усилили тем самым его мозг ручкой, бумажкой и методом вычисления, может оказаться в этом умножении много искусней человека с высоким IQ, который пытается умножать в уме, да ещё и незнаком с методом умножения в столбик. Что уж говорить о более сложных умениях! Человек с низким IQ, который вовремя догадывается задавать вопросы Гуглу, легко победит в интеллекте человека с высоким IQ, который игнорирует Гугл и пытается каждый раз что-то сообразить самостоятельно. При этом лучшие умы человечества как раз используют Гугл, не стесняясь⁸³ – а их «аппаратный IQ» вообще перестаёт играть большую роль. Большую роль играет образование (включая умение пользоваться инструментами!), использование этих инструментов (не лениться использовать! Всё записывать!), а также решение цепочек всё более трудных проблем вместо безуспешного решения за пределами трудной проблемы «в лоб» (важны способы решения проблем, которые используются, и этому тоже надо учиться).

⁸¹ https://en.wikipedia.org/wiki/Intelligence_quotient

⁸² пре-принт <https://www.medrxiv.org/content/10.1101/2021.08.10.21261846v1.full.pdf>

⁸³ «Этот человек в топ-1% по навыкам кодинга, скорее даже 0.01%, и при этом он делает то, чего я часто не вижу у вопрошающих людей в Slack/телеграм чатах/у стажеров на работе. Он видит ошибку И СРАЗУ ЕЕ ГУГЛИТ, НЕ СТЕСНЯЯСЬ. И, внешне, находит решение. Я не знаю, почему люди этого не делают, быть может думают, что умные и сами разберутся» – <https://t.me/seeallochnaya/414>

Культурная (следующая каким-то выработанным цивилизацией шаблонам быстрого эффективного мышления, а эти шаблоны даёт хорошее образование) работа побеждает «творческое дикарство», равно как использование дополнительной вычислительной аппаратуры (от ручки-бумажки до современных дата-центров с софтом AI) уравнивает жизненные шансы людей с низким IQ и высоким IQ.

Кейс: интеллект как способность научиться

Собачка Тяпа научилась бегать и прыгать за полгода, а девочка Таня за пару лет, и то не очень. Какие из этих высказываний верны:

— У Тяпы интеллект сильнее, ибо она научилась бегать и прыгать быстрее, чем Таня. ### Нет, интеллект сильнее у Тани, ибо она научится не только бегать и прыгать, но ещё и читать-писать, готовить, программировать, играть на рояле, а собачка Тяпа – никогда. Тут речь идёт об интеллекте в целом, его сила определяется по уровню общности в решении задач.

— У Тяпы телесный интеллект сильнее, ибо она научилась бегать и прыгать быстрее, чем Таня. ### Нет, телесный интеллект сильнее у Тани, ибо она быстро научится не только бегать и прыгать, как собачка, но ещё и петь (мышечная работа), рисовать (мышечная работа), танцевать (мышечная работа), поднимать штангу (мышечная работа). А собачка Тяпа всего этого не сможет.

Пётр – отличный математик, никогда ничем кроме математики не занимался, «узкий специалист», но лауреат какой-то важной математической премии. Павел – занимался самым разным, и даже немножко математикой (совсем чуть-чуть, это оказалось «не его»), но там была и инженерия, и исследования по химии, и много чего ещё. Разбираться приходилось с нуля в каждом новом проекте. Премий Павел не получал, но со всеми работами справлялся, и часто получше, чем местные «узкие специалисты». У кого интеллект сильнее?

— У Петра, у него ведь математическая премия ### Нет, интеллект определяется как эффективность в обучении новому, причём акцент на разнообразии этого нового. Пётр узкий специалист, у него интеллект «широкая генерализация» (мастерство), но не «экстремальная генерализация» (на что претендует Павел).

— у Петра, он настоящий, глубокий специалист! ### Нет, «глубокий специалист» указывает как раз на узость предметной области (хотя сама по себе математика не так уж и узка, как предметная область, но в разнообразии задач, стоящих перед людьми, математические задачи – это крошечный класс задач)

+++ У Павла, он быстро научится любому новому делу ### Да, сила интеллекта определяется именно этим: скоростью обучения новым делам, акцент на широте этих дел, их разнообразии. «Любое новое дело» указывает на «экстремальную генерализацию» (то, чему можно научить людей), хотя понятно, что это сильное преувеличение, бытовой речевой оборот. Тем не менее, в определении силы интеллекта разнообразие выигрывает у узкой специализации.

Богдан на выходе из средней школы имел IQ 120, Марина IQ 130, Дженнифер IQ 140.

??? Кто из них разбогател через 15 лет после окончания школы?

+++ нельзя сказать ### Да, это невозможно сказать, ибо богатство больше определяется случаем, чем целенаправленными усилиями. IQ тут не влияет. Кроме того, мы даже не можем сказать, кто из них стал за 15 лет умнее (ибо кто-то мог получить хорошее высшее образование, а кто-то так и остался неучем – и никакой IQ тут не поможет). С другой стороны, статистика говорит, что у Дженнифер в плане «разбогатеть» шансы больше!

— разбогатеет Дженнифер, у неё IQ выше всех ### нет, богатство и IQ не непосредственно связаны друг с другом. Во-первых, интеллект можно поднимать образованием, а во-вторых, богатство больше зависит от случая – чтобы стать богатым, не нужно иметь сверх-высокий IQ, достаточно не делать грубых ошибок, и чтобы удача была на твоей стороне. Поэтому нельзя сказать, кто разбогатеет. Может быть, и Дженнифер, но это определяется отнюдь не только её высоким IQ.

— разбогатеет Марина, у неё средний IQ, поэтому у неё всё сбалансировано ### Нет, нельзя сказать, кто разбогатеет. Богатство практически не зависит от IQ, а «всё сбалансировано» – это вообще непонятно, что имеется в виду.

??? Кто из них стал самым успешным учёным, если известно, что все трое пошли в науку?
+++ нельзя сказать ### Да, это невозможно сказать, ибо успех в науке больше определяется случаем, упорством, организованностью, задействованием компьютеров, но не IQ.

— лучше всех станет Дженнифер, у неё IQ выше всех ### нет, нобелевские премии получают учёные не с самым большим IQ, равно как разброс по IQ среди именитых и продуктивных учёных весьма велик. Может быть, великим учёным станет именно Дженнифер, но не по причине её высокого IQ.

Мышление – это функция/ поведение/назначение интеллекта

Мышление – это поведение интеллекта, его функция. Интеллекты бывают разной направленности (удачные для разных классов проблем, которые только можно представить во вселенной – помним, что вычислители неодинаково эффективны для разных классов вычислений, теорема об отсутствии бесплатного обеда), разного калибра/силы/уровня/общности/эффективности в части «отращивания» разных видов прикладного мастерства. Учим интеллектам разной направленности и силы – учим мышлению разной направленности и силы. Качество мышления обученного нами интеллекта мы должны смотреть не на знакомых ему в ходе обучения ситуациях, и даже не на знакомых нам, его учителям, ситуациях, а на незнакомых ситуациях – на решении проблем, которые ранее ещё не встречались. И не в условиях «экзамена», а в условиях реальной жизни, в реальных проектах. Интеллект – это когда ты изучаешь что-то новое, научаешься новым мыслительным операциям, которые потом войдут в прикладное мастерство.

А что же с мышлением в ходе решения прикладных задач? Если будут затыки/проблемы, то это будет мышление. Если просто вы ещё и ещё раз будете решать знакомую вам задачу, то это будет не мышление. **Мышление – это когда один алгоритм-интеллект составляет другой алгоритм-объяснение, кодирует правила рассуждений в объяснениях для незнакомой ранее предметной области. Если же просто производится работа прикладного вычислителя-мастерства, то в нашем случае это прикладные рассуждения, работа уже выученного робота, автоматизм.** Конечно, это очень условное разделение, но оно кажется полезным, если обсуждать, каким образом прирастают знания и умения агентов. Прирост знаний – результат мышления, которое с учётом выхода «вычислений» в физический мир (эксперименты) называют «познание», а в машинном интеллекте предпочитают называть learning.

Конечно, мышление включает «просто рассуждения»/inference! Без этого никак! Это всё вычисления как операции над изменениями информации, записанной в памяти, причём эти операции делаются по определённым «правилам вывода/рассуждений/inference» – это и есть «рассуждения»/inference. Но вот использование знаний, полученных мышлением – это «просто рассуждения», а не «рассуждения мышления». Нам просто удобно разделить рассуждения: вычисления на происходящие при мышлении интеллекта и происходящие при пользовании прикладным мастерством. Так что интеллект можно задействовать для улучшения не только прикладного мастерства, но и рассуждений самого интеллекта, отрастить себе новую версию какого-то мыслительного мастерства, или даже отрастить её не в себе, а в инструменте, например, компьютере – или даже в нанятой для этих рассуждений фирме! И всё это оперирование с практиками требует интеллекта.

Если вы умеете читать, то вы просто читаете, задействуете привычное мастерство чтения, а не мыслите про чтение. Если вы умеете считать, то вы просто считаете. Работа интеллекта, мышление нужно было, когда вы знакомились с чтением и письмом, осваивали эти дисциплины. Мышление у вас работает в вузе, когда вам нужно разобраться за пару месяцев с очередной парой толстых томов с формулами. А когда вы уже пятый год на работе просто применяете эти формулы, вы это делаете автоматически, мышления не происходит – пока вы не встретитесь с проблемой, которой раньше не было. Только в этот момент вы включаете мозг, ту его часть, которая ответственна за интеллект. И эта часть начинает работать – эта работа и есть мышление. Если проблем долго нет, то мозг пластичен: мышление не включается, пластичный мозг потихоньку деградирует, сила интеллекта потихоньку падает. В текущем году это падение с лихвой компенсировано информационно-коммуникационными технологиями: раньше

нужно было «придумать решение проблемы», сегодня нужно «не забыть погуглить решение проблемы». Проще простого перейти в режим неинтеллектуальной обезьянки, которая проблемы не решает, но бодро щёлкает задачки, на которые она была адрессирована раньше – и так живёт годами, пока не окажется, что интеллект совсем зачах, прошивка мозга устарела, жизнь несётся мимо, и непонятно как вернуть те времена, когда интеллект в ходе обучения и решения проблем непрерывно усиливался, а не деградировал. Интеллект должен расти всю жизнь, это не дело, когда мышлением люди занимаются последний раз в вузе!

Напомним, что поведение вычислителя определяется не только и даже в силу универсальности вычислителей, не столько аппаратурой (хотя скорость работы аппаратуры и физика в основе работы аппаратуры – биологические нейроны, классическая электроника, квантовые явления влияют на поведение вычислителя), сколько программным обеспечением, «софтом». Тезис Тьюринга-Чёрча-Дойча про универсальность вычислителя говорит, что все вычислители независимо от физической их природы умеют вычислять ровно столько же видов функций, сколько простейшая машина Тьюринга, просто скорость вычисления будет разная. Этот тезис подробно раскрывается Дэвидом Дойчем в его книжках. И вообще, граница между аппаратурой и софтом весьма размыта.

Мы это для случая интеллекта-вычислителя и мышления как его вычислений формулируем так, что интеллект может быть не только врождённый «аппаратный» (человеческий, машинный, человеко-машинный, коллективный для людей и машин как аппаратных вычислителей, пришедших «с завода», без «предустановленного софта», необразованных), но и выученный/learned. И машины, и люди, и даже коллективы должны быть обучены, чтобы в них появился «софт» алгоритмов сильного интеллекта. Врождённого интеллекта никогда не хватает!

Можно говорить как об усилении интеллекта (вычислитель как функциональный объект), так и об усилении мышления (поведение вычислителя, его функция) – по сути, это одно и то же. «Мышление» неуловимо, как и любое поведение/работа: процессы сложно представлять, их сложно обсуждать. А интеллект как функциональная часть мозга, ответственный за освоение нового мастерства – вполне понятно, как о нём думать. В нужный момент, при появлении новой задачи, он включается, и начинает мыслить, то есть мастерить другую функциональную часть мозга, которая называется «прикладное мастерство» и будет ответственна за рассуждения по решению «на автомате» какого-то класса прикладных задач. Или даже какое-то мастерство (например, в логике) может быть ответственно за решение «на автомате» задач самого интеллекта! Поэтому развиваем интеллект (в инженерии было бы «создаём и развиваем», но мы не создаём врождённый интеллект в людях, а только развиваем его. Но в случае AI мы этот интеллект ещё и создаём), а уже потом развитый/усиленный интеллект проявляет сильное мышление во время его использования.

Всё, конечно, не так просто. Мы говорим про функциональную часть мозга, или функциональную часть компьютера, или функциональную часть мозга и компьютера вместе (гибридный интеллект человека и компьютера), или функциональную часть мозгов и компьютеров группы людей, да ещё и с неизвестным сегодня науке способом реализации конструктивными частями – анатомическими структурами мозга. Разве что в случае компьютеров тут можно рассказывать, как именно происходит мышление или рассуждения, но и тут есть оговорки: если речь идёт о компьютерных нейронных сетях, то до сих пор не очень понятно, как именно они работают. Но главное: никакой мистики, никакой психологии! Мышление, интеллект, правильные рассуждения, мастерство – обо всём этом мы можем говорить инженерно, и включать в рассуждения не голого человека, и обязательно одного, а команды людей с их компьютерами. И обсуждать как познание (обучение и исследования, образование и науку), так и работу с достижением целей на основе уже познанного: приложение мастерства.

Важно различать в обучении то, что ведёт к усилению интеллекта (знаний о том, как получать знания, как решать проблемы) и что ведёт к увеличению объёма прикладных знаний (знания о том, как решать какие-то известные классы задач, например, «умножать столбиком, если под рукой нет калькулятора, хотя сегодня калькулятор под рукой есть всегда»). При всей условности различения этих знаний, для образования нужно приоритетно выбирать знания по усилению интеллекта, то есть знания практик интеллект-стека: трансдисциплины интеллект-стека и инструменты (прежде всего моделеры) для поддержки рассуждений по этим трансдисциплинам. Это позволит решать всё более и более сложные проблемы, в том числе проблемы, связанные с развитием практик самого интеллект-стека.

Чему учиться уже образованным?

Занимаясь парой-тройкой прикладных практик, умнее не станешь, интеллект не разовьёшь. Для усиления интеллекта надо заниматься трансдисциплинами: сначала просто освоить лучшие их версии, известные на сегодняшний день, а потом пытаться решать проблемы создания новых версий трансдисциплин интеллект-стека, которые будут лучше сегодняшних (то есть заниматься мышлением по поводу самого мышления).

Трансдисциплины могут быть неосознаваемые «народные», «самопальные» («здравый смысл», а не математическая логика), или наоборот – хорошо осознаваемые лучшие известные на данный момент человечеству, SoTA. Много ли таких SoTA трансдисциплин вы изучали в школе, бакалавриате, магистратуре традиционной государственной системы образования? Можно поспорить, что ничтожное количество. Вот физкультура там была предметом, который понятным образом влияет на качество последующей жизни: здоровое тело может долго поддерживать ясность внимания, меньше уставать за полный рабочий день. Но даже физкультура (главным образом командные игры: баскетбол, волейбол, и немного лёгкая атлетика) не подавалась для этих целей. А для чего? А непонятно для чего! Для сдачи норм ГТО («готов к труду и обороне»), рудимент эпохи примата физического труда и милитаристской организации общества.

Большинство других предметов имели более чем прикладное значение (даже физика и математика!), сегодня их знание не помогает ориентироваться в непрерывно меняющейся жизни, не используется никак. Когда вы в последний раз задействовали знание различия дифракции и интерференции из курса оптики или закона Кирхгофа из раздела «Электричество» школьного курса физики? Сходу можете сказать, чем отличается момент инерции и импульс? А это вы всё учили как «базовые знания, которые пригодятся в жизни!» Ну что,годились ли в жизни, илигодились только при сдаче экзамена и при изучении вузовского курса физики, который так же в жизни никак не пригодился? А учили ли вас логике как искусству правильных рассуждений, и если таки случайно учили, то сколько времени от времени всей школьной и вузовской программы? И какой версии логики вас учили? Аристотелевская логика ведь давно была «уволена» примерно так же, как была уволена теория флогистона и алхимия: она плохо работала! Нашлись варианты логики получше, state-of-the-art.

Прошивку интеллекта, полученную «исподволь» (не прямым обучением трансдисциплинам, а путём накопления более-менее случайного опыта при изучении каких-то прикладных дисциплин) в традиционном образовании, нужно менять на современную текущего года, нацеленную на будущую жизнь в условиях полной рабочей неопределённости. Помним, что интеллект работает в условиях, о которых не догадывается ни ученик, ни его учитель.

Вспомним игровую метафору. Мышление как работа интеллекта нужно, чтобы научиться играть разные проектные роли в проектах, как в ролевых играх. Мастерство в каких трансдисциплинах, исполнение каких ролей даст нам мастерство в мышлении, то есть мастерство справляться со всё новыми и новыми ситуациями, с новыми и новыми задачами? Ведь каждый раз, когда нам нужно зайти в проект, интеллект должен выбрать подходящее мастерство, сориентировать агента на занятие роли, далее следить, чтобы не было проблем. То есть мы видим агента-киборга внутри проекта в какой-то роли



Мыслительные практики (практики, которые исполняет интеллект) тоже имеют названия ролей, которые занимает агент, чей интеллект выполняет эти практики. И эти практики работают с какими-то функциональными объектами: суть практики в том, чтобы выделять вниманием из пёстрого и мелькающего окружающего и виртуального/ментального/абстрактного миров объекты и проводить с ними какие-то рассуждения по правилам.

Практика (по имени транс- дисциплины)	Роль	Некоторые функциональные объекты практики
Системная инженерия	Инженер	Практики и их роли: разработки, архитектуры, DevOps/platform engineering. Концепция использования, концепция системы, архитектура и модульность, инженерные обоснования и прохождение развилки, постоянный ввод в эксплуатацию. Мемом и феном в техноэволюции. Архитектурные идеи Дойля.
Методология	Методолог	Практика/деятельность/рабочий процесс создателя/труд, роль/функция агента-создателя, жизненный цикл и бесконечное развитие. Агент, прагматизм (цели и средства), блага, стратегия, альтернатива, предпочтение, (в том числе рациональный) выбор, ресурс, субъективная полезность, убывающая предельная полезность ресурса, отрицательная полезность работы, межвременные предпочтения, альтернативные издержки, предельные издержки, маржинальный выбор.
Риторика	Ритор	Убедительное объяснение, модель знаний, эмоций и др. агента, понятийное расстояние, нарративизация
Этика	Совесьть	Этические принципы, оценка целей, оценка средств. Чья этика (проблема индивида и конфликты системных уровней).
Эстетика	Эстет	Красота, элегантность
Познание	Исследователь	Недообъяснённый участок мира (проблема), догадка, критический рационализм (опровергаемость без доказательств), дополнения Дойча: трудноварьируемость объяснения, универсальность, компактность, оптимизм, бесконечность познания, структура реальности (фундаментальные объяснительные теории). Творчество и новизна. Свобода

Практика (по имени транс- дисциплины)	Роль	Некоторые функциональные объекты практики
		критики как условие познания. Явления типа социального лазинга и меметика в целом.
Рациональ- ность	Разум	Описания мира (модели) и рациональность против эмпирицизма (в том числе дискуссия про достаточность RL и эволюцию с free energy как отсутствие глобального “подкрепления”, функции наград временные и познаются/выучиваются). Теория решений как выбор из альтернатив: causal quantum-like active inference (excess Bayesian inference). Три уровня лестницы понимания/объяснения/причинности (ассоциации, “если это, то что” – интервенции, “если бы не” – возможные миры и контрфактуалы). Действие как реакция на изменения в окружении (active/embodied, а не любой inference).
Логика	Логик	Рассуждение по правилам (с заданными типами по заданным алгоритмам), правила рассуждений (булевых, байесовских, квантовоподобных), аргументы, ошибки в рассуждениях и отношение к ним («а чо такова?»). Формальная несовместимость разных онтологий, микротеоории/онтики.
Алгоритмика	Алгоритмист	Вычисления (включая понятие inference и update) и вычислимость теоретическая и практическая, интеллект, физичность и универсальность вычислителей (включая универсальный аппроксиматор – нейросеть), сложность вычисления, алгоритмы, способы описания вычислений/алгоритмов (парадигмы программирования, включая вероятностное и квантовое). No free lunch theorem, проблема равенства P и NP.

Практика (по имени транс- дисциплины)	Роль	Некоторые функциональные объекты практики
Онтология	Онтолог	абстрагирование/моделирование и рендеринг/конкретизация/демо моделирование, уровни абстракции: метамоделирование (foundational, upper, middle, work and operational ontology как M4-M0). Вероятность в различении типов.
Теория понятий	Типолог	типы, классы, отношения композиции, отношения классификации, отношения специализации, отношение передачи/ изготовления (кто что кому даёт: трёхместное).
Физика	Физик	Физический реализм, безмасштабность. Термодинамика (действие, энергия, свободная энергия, фазовое пространство), теория информации (в том числе суперинформации и в том числе “неизменности репликации”, в том числе термодинамическая семантика), физические принципы и законы. Квантовость, эргодичность. Стохастика: эмерджентность и фрустрации , ренормализация. Негэнтропия, эволюция как обучение и цель жизни (плюс центральная догма молекулярной биологии). Создатель (constructor theory) и введение объектов стека создания.
Математика	Математик	Множество, аксиома, алгебра/исчисление, пространство, поле, счётная и несчётная бесконечность. Субъективная теория вероятности (нефизичность вероятности), колмогоровская и неколмогоровская вероятности, категория.
Семантика	Семантик	Объекты, отношения, имена, язык, значения, физичность всего происходящего, описание, отношение репрезентации, математический объект, математическая операция, уровень формальности (строгость типизации) и понятие “доказательства”.

Практика (по имени транс- дисциплины)	Роль	Некоторые функциональные объекты практики
Собранность	Собранный (киборг!)	Собранность/self-collectedness, сознание как mind control, внимание, экзокортекс, тело и сома , ловкость/body control. Безмасштабность собранности, понятийное наведение и удержание внимания.
Понятизация	Поэт	Фон, прототип, образец, паттерн/фигура (свойства: похожесть, вероятность отнесения к категории и т д), ритм/последовательность, стиль, модальность восприятия, неудовлетворенность/проблема . Мир и «я», деятельное «ощупывание».

Конечно, трансдисциплины интеллект-стека в любой его версии представляют собой плотно сплетённую сеть («клубок») объяснений, в которых они тянут какие-то тематические нити, плотно спутанные между собой. Примерно так об этом говорит Дэвид Дойч в книге «Структура реальности», он там выделяет четыре объяснительные нити, которые он считает самыми важными как лежащие в основе всех других объяснений, да ещё и переплетёнными так, что объяснения каждой из них невозможны без объяснений других нитей:

- Квантовая физика. Дойч считает, что фронтир тут – в интерпретации многих миров Эверетта.
- Эволюционная эпистемология и критический рационализм Поппера
- Вычисления и универсальный компьютер Тьюринга
- Меметика и эволюция мемов Докинза

В наш вариант интеллект-стека всё это вошло, но мы не просто перечисляем какие-то объяснения. Мы, как и Дойч, говорим, что эти объяснения дают возможность бесконечного познания, в том числе бесконечного усиления интеллекта, но мы ещё и говорим, что этим объяснениям надо целенаправленно учить и людей, и AI. Поэтому у нас есть задача нарративизации, то есть последовательного изложения этого клубка идей в развёрнутом тексте учебных курсов. Мы весьма условно растягиваем клубок объяснений разных трансдисциплин на отдельные части и располагаем их весьма условно в виде стека. Мы просто очень грубо оценили, что трансдисциплины/объяснения верхних уровней стека используют трансдисциплины/объяснения более низких уровней больше, чем наоборот (помним, что это плотно перепутанная сеть объяснений!).

Мыслительное мастерство поэтому нельзя приобрести, если просто «выучить всё снизу вверх». Нет, поскольку там клубок, то для того, чтобы последовательное изложение дисциплин как-то собралось в голове в связную картину мира, потребуется специальная организация учебного курса. Упоминание понятий, которые ещё не объяснены, неизбежно – и поэтому либо потребуется дважды проходить короткую последовательность курсов, чтобы откорректировать на втором проходе непонимание первого прохода, или иметь длинную якобы «однократную» версию с неизбежными повторами.

Жизненное мастерство агента в целом как мастерство отличной жизни (в том числе жизни компании! **Жизнь – это просто поведение агента в ходе его существования**), включает:

- непрерывно развиваемое **прикладное мастерство** («компетенции» для рынка, в том числе для рынка труда, если речь идёт о людях, чтобы мочь получать ресурсы для жизни), и
- **мыслительное мастерство/интеллект**, как врождённый (который не надо учить, доступный в момент создания агента), так и познанный. Если речь идёт о человеке, то это получаемый от родителей (гены врождённого интеллекта) и в его научаемой части в детском саду, школе, бакалавриате, в семье, в кружках, самообразованием. **Мы считаем, что это владение трансдисциплинами из интеллект-стека на некотором уровне беглости, причём и личное владение, и командное владение, в том числе с учётом доступа к AI и интернету.**

План действий по приобретению жизненного мастерства (очень условно, но речь идёт и о людях, и о компьютерных AI, и даже о фирмах с их гибридным коллективным интеллектом, это рассуждение общее для агентов любой природы):

1. Сначала нужно обучиться сильному мышлению, то есть поднять силу своего интеллекта, стать умным (если это фирма, то нанять умную команду и AI поумней, или нанять любую команду, но затем дать ей фундаментальное образование).

2. Затем нужно научиться каким-то прикладным практикам, за выполнение которых будут платить деньги в проектах. Прикладные практики дадут деньги, на которые можно будет продолжать что-то делать. Что делать? Избегать неприятных сюрпризов, в том числе на уровне человечества в целом (скажем, принимать участие в организации политической жизни без войн, добиться решения проблемы биологического бессмертия, решить проблему ядерного синтеза для получения больших количеств энергии, космического расселения для предотвращения рисков столкновения цивилизации с астероидами, и т.д.).

Выбор того, в какое мыслительное или прикладное мастерство в каких количествах инвестировать своё учебное время, делается практикой стратегирования (она описана в курсе «Системный менеджмент», это практика общая для стратегирования агента-личности и агента-фирмы).

За силу интеллекта (мастерство в мышлении, мыслительное мастерство) не платят, платят за приложение прикладного мастерства, которое с интеллектом связано только тем, что сильный интеллект позволяет его приобрести много быстрее. Прикладное мастерство придётся менять довольно часто, ибо практики устаревают. Нельзя ожидать ни от людей, ни от компьютеров, ни даже от фирм, что выучился какой-то профессии – и это на всю жизнь. Нет, новым прикладным практикам надо учиться всю жизнь, а для этого нужен сильный интеллект.

Если бросить все силы только на получение прикладного мастерства, чтобы стать в нём лучшим в мире, и проигнорировать фундаментальное образование (то есть дисциплины интеллект-стека), то жизнь в целом лучше не станет: менять проекты (оказываться в новой предметной и организационной ситуации), и в связи с этой сменой проектов прикладное мастерство (исполняемую в проекте основную роль) придётся довольно часто, ибо мир не стоит на месте, так что быть прикладным мастером без мастерства в мышлении (то есть без сильного интеллекта) не получится. У глупых людей (людей с низким интеллектом) возможности в жизни сильно ограничены, равно как и у глупых фирм!

Ситуация осложняется тем, что практики интеллект-стека тоже довольно быстро меняются, прогресс не стоит на месте. Но ситуация облегчается тем, что можно чередовать образование по фундаментальным и прикладным практикам, делать «блинчатый пирог» в своём

образовании. Но в любом случае, надо делать это образование непрерывным, заниматься им всю жизнь, развиваться, приобретая новые «фичи», то есть научаясь выполнять всё новые и новые практики. Это относится к людям, к AI, к коллективам людей и AI (организациям), сообществам, человечеству в целом.

Так что получаем сначала мастерство в мышлении (сильный интеллект на базе освоения практик интеллект-стека), а затем используем его по прямому назначению: разбираемся во всё новых и новых проблемах (и быстро обучаясь, если знания уже где-то есть, или самостоятельно проводя исследования, если знаний ни у кого на планете нет). Это даёт возможность исполнять всё новые и новые роли (их жизнь будет подкидывать с избытком), решать всё более и более трудные проблемы, причём за деньги. **Бесконечно развиваемся при помощи интеллекта, наносим непоправимую пользу человечеству прикладным мастерством, и нам за это платят! Это и есть жизнь, задействование жизненного мастерства!**

Интеллект, конечно, может быть использован и для усиления самого себя, мастерством собственного мышления можно поднимать своё мастерство мышления ещё выше. Каждый агент сам решает, сколько времени «мозговые мышцы» качать, развивая мозг с экзокортексом в части трансдисциплин, а сколько времени использовать текущую силу интеллекта для прикладной работы в проектах. Хотите небольшие результаты прямо сейчас, или большие немного погодя? Синицу в руках прямо сейчас, или журавля в руках – но попозже? Да, это вечный вопрос выбора между познанием и использованием знания, *exploration* против *exploitation*, классическая проблема, не имеющая математического решения⁸⁴. И это вопрос как для людей, так и для AI, так и для компаний.

⁸⁴ <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095693>

Упражнение: оценка мыслительного мастерства

Оцените по десятибалльной шкале, насколько вас научили мыслительному мастерству в вузе, и сколько вы добавили самообразованием (в сумме 10 баллов на вуз+самостоятельное изучение). Это ничего, что вы не очень понимаете пока содержание трансдисциплин интеллект-стека: попробуйте догадаться по тем функциональным объектам, которые являются предметом этих трансдисциплин. Дальше в курсе будут приведены краткие пояснения по каждой трансдисциплине, но сначала попробуйте оценить, насколько вы владеете культурой современного мышления, практиками интеллект-стека.

Практика интеллект-стека	Оценка текущего вашего мастерства в практике по десятибалльной шкале	Из них научили в вузе	Из них научился сам
Системная инженерия			
Методология			
Риторика			
Этика			
Эстетика			
Познание/исследования			
Рациональность			
Логика			
Алгоритмика			
Онтология			
Теория понятий			
Физика			
Математика			
Семантика			
Собранность			
Понятизация			

Особую пикантность моменту придаёт то, что SoTA практик интеллекта как мыслительного мастерства кардинально поменялось уже в 21 веке. Хорошее фундаментальное образование прошлого века работает сегодня примерно так же хорошо, как в физике работала теория флогистона до прихода современной термодинамики. То есть это образование 20 века вроде как работает, но по современным критериям очевидно, что работает плохо! Не может образование прошлого века сегодня работать хорошо! Фундаментальное знание изменилось! Теория флогистона когда-то отлично работала, потому что никто не знал, что можно вычислять точнее. Затем теория флогистона была заменена в физике более современными теориями, которые постепенно (не сразу!) попали и в массовое образование. А вот набор практик интеллект-стека в их современном состоянии пока в массовое образование не попали. Люди по-прежнему думают, что аристотелевская логика и есть та логика, которой нужно пользоваться.

Но нет, математическая логика давно заменила аристотелевскую логику, в логике аристотелевская логика имеет примерно такой же статус, как теория флогистона: относится к истории логики, а не к современной логике, как и теория флогистона относится к истории физики, а не современной физике.

И не забываем, что после того, как интеллект отработал и вы сориентировались в ситуации, вам нужно будет задействовать своё прикладное мастерство: деньги-то вам заплатят за выполнение работы, а не за хорошее понимание ситуации. Если вы (или ваш компьютер с AI, или ваша фирма) торгуете хорошим пониманием ситуации, то вам придётся достичь и в этом уровня прикладного мастерства, иначе проиграете конкурентам!

#

Интеллект против культа карго

Интеллект работает на простом принципе: он фильтрует многообразие окружающего мира, концентрируя внимание на определённых его объектах и отношениях (или в случае перехода на конструктивные онтологии – объектах и операциях с ними). Интеллект даёт чеклист – что надо заметить в мире и его моделях, на что обратить внимание, о чём не забыть подумать, а что наоборот – откинуть. Только самое важное, только самое надёжное. Все трансдисциплины – это такие чеклисты для определения самого важного и надёжного в самых разных ситуациях.

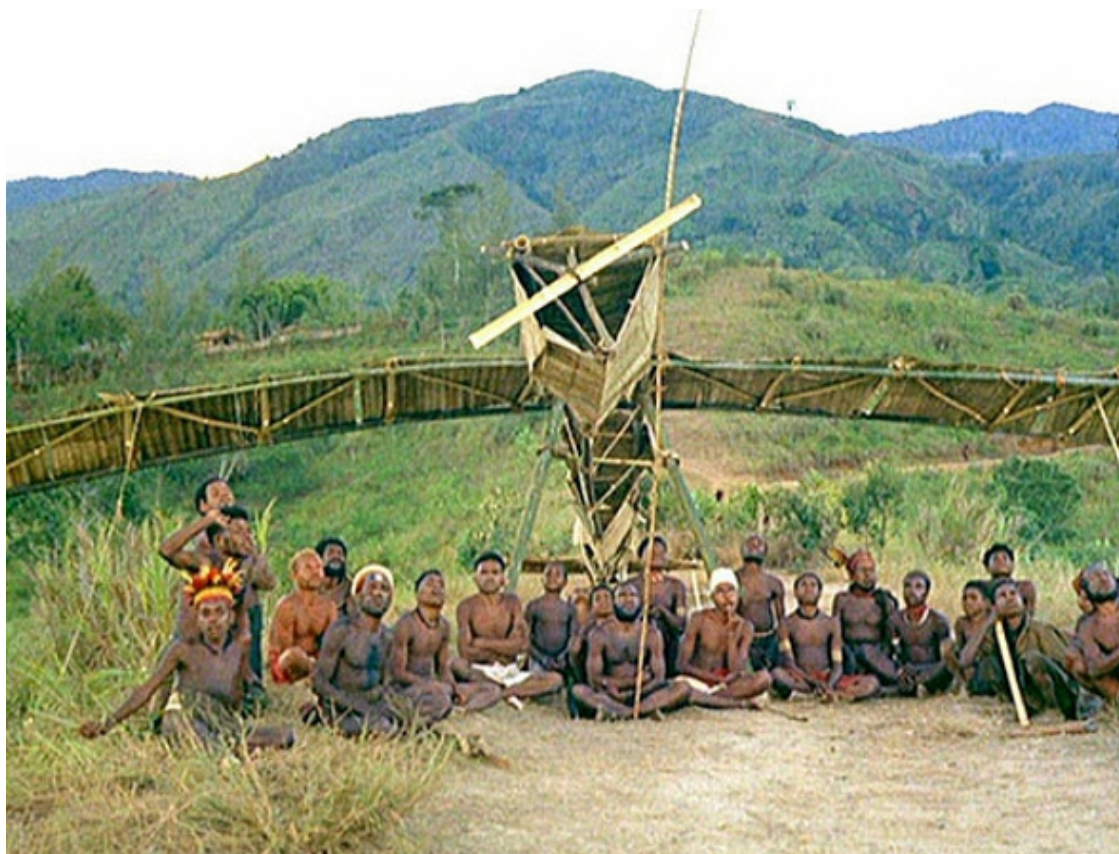
Вы заходите в комнату с работающими людьми, вы хотите разобраться в том, что они делают. Системное мышление подсказывает: эти люди занимаются какой-то системой, выясни у них, какой. Это и есть «проект» по созданию и развитию какой-то системы. Ты выясняешь – ибо если это атомная электростанция, то это один разговор, а если вечер фортепианной музыки – то разговор будет совершенно другой. Все эти люди в проекте по поводу этой системы играют какие-то роли и выполняют какие-то практики, и вы быстро ориентируетесь, с кем о чём разговаривать и как их заинтересовать в ваших проблемах. Методология позволяет внятно описать саму деятельность, а не мычать что-то на эту тему. Системное мышление подсказывает: они тут сидят в том числе и потому, что какие-то внешние по отношению к этому проекту люди играют какие-то внешние проектные роли. Вы интересуетесь этим, и становится понятно, зачем эти собравшиеся в комнате люди собрались, что они делают, вы можете как-то предсказать результат проекта и принять решение о вашем дальнейшем участии.

Если ситуация так и осталась непонятной, то вы начинаете выдвигать какие-то догадки и проверять их. Это познание/исследование: вы абсолютно осознанно выдвигаете эти догадки/гипотезы, осознанно их проверяете. Вы не путаете физический мир и описания, знание онтологии в том числе и про это. Иногда даже начинаете заниматься созданием новой прикладной дисциплины (ибо в совсем новых проектах может встретиться деятельность, которой раньше не было), задействуя исследование и методологию.

При этом вам хватает остроты внимания на полный рабочий день, это своё внимание и память вы поддерживаете записями в компьютере – это даёт мастерство собранности. Иногда вы строите какие-то модели ситуации (или задействуете модели сидящих в комнате ваших собеседников), компьютер проводит над ними вычисления – это вы используете мастерство моделирования, включающее в себя и мастерство понятизации, и семантики, и теории понятий, и логики. А ещё вы можете объяснить другим людям ваше понимание, это даст рациональность. Сможете убедить их сделать что-то полезное для улучшения ситуации – это даст риторика.

Понятия каждой упомянутой трансдисциплины делают именно это: заставляют о чём-то думать, а что-то из мышления выкидывать в силу неважности, экономить мышление. Если у вас нет современной версии интеллекта, который даёт вам лучшие на сегодня способы мышления, то весьма вероятно, что вы создадите в проекте мыслительный культ карго⁸⁵.

⁸⁵ <https://ru.wikipedia.org/wiki/Карго-культ>



Во время второй мировой войны в Меланезии было замечено, что дикари строят буквально из коры и веток модели самолётов. Почему? Потому что на самолётах прилетали посланники богов, и давали им дары богов: консервы, одежду (cargo/груз на самолётах). Так что они обращали внимание на самое важное: больших птиц из непонятных материалов. Они делали таких птиц, и ожидали, что это привлечёт ушедших после окончания войны посланников богов, и они появятся из этих птиц и опять одарят их разными полезными диковинами.

Если вы не имеете понятийного мышления, привлекающего внимание к действительно важным объектам (об их важности известно из опыта человечества, данного в виде лучших на сегодняшний день объяснений/теорий/трансдисциплин), то вы в незнакомой вам ситуации неизбежно создадите прикладной карго-культ! Соорудите себе мышление из коры и веток, но оно не сможет летать.

Интеллект для того и нужен, чтобы вы смотрели не глазами дикаря, а глазами современного человека, вооружённого самым изощрённым мышлением, которое придумала человеческая цивилизация на настоящий момент.

Дикарский мир 21 века: переучиваться самому, жалеть и учить других

Дикарей нужно жалеть и учить. Население глобуса в его большинстве (кто думает только о какой-то одной стране, тот думает местечково) автор объявляет агрессивно и неполиткорректно дикарями. Почти всё это население училось в школе. Огромное число этих дикарей училось ещё и в вузах, а некоторая часть представляет собой седовласых профессоров или даже академиков. Всех их надо жалеть и учить, жалеть и учить, причём и профессоров, и академиков тоже (хотя в каких-то очень узких предметных областях профессор и академик могут быть вполне на фронтире, но мы помним, что прикладное даже фронтирное мастерство – это ещё не интеллект, интеллект занимается как раз новым и неизведанным и его свойством является универсальность/широта, а не узость и прикладность).

Ликвидация безграмотности осталась кампанией уже вековой давности, когда за парты садились и стар и млад, когда читать-писать никто не умел. А сейчас читать-писать-считать все умеют. Но **лучшие известные цивилизации способы мышления изменились существенно в 21 веке, и культурные и передовые профессора двадцатилетней давности оказались дикарями в 21 веке точно так же, как дикарями в начале 20 века оказались преподаватели теорий витализма и флогистона.** Изменилось всё, даже сам 21 век уже обычно не пишут римскими цифрами, так его писали только старики, пришедшие из XX века.

Автор лично попал в группу студентов, которые были вынуждены сдавать два экзамена по квантовой химии: один по с трудом уцелевшей в СССР, но потом всё-таки списанной в утиль по старости теории резонанса⁸⁶, а другой по относительно молодой тогда (конец 70-х двадцатого века) теории молекулярных орбиталей⁸⁷. Это был последний год, когда теория резонанса преподавалась в вузах, и нам было объяснено, что такова традиция, и её сходу не поменяешь: учителей быстро не переучишь, и нам просто не свезло: сдавать экзамен на знание выкинутой на свалку научной истории дисциплины надо по не спрашивайте каким соображениям – социальным, историческим, административным, но ни разу не научным, не рациональным. Это было ещё в прошлом веке, теории сменяли друг друга не спеша, не было интернета, не было свободного доступа к научным журналам через планшет в парке на скамеечке.

Сегодня ровно такое массовое преподавание дисциплин со свалки научной истории происходит не с узкими прикладными теориями, но с фундаментальными, лежащими в основе светского/научного мировоззрения трансдисциплинами.

Современные «мировоззренческие» теории оказываются зубодробительными для почтенной профессорской публики примерно как сама идея современной экспериментальной науки была зубодробительна для тогдашней почтенной публики где-нибудь во времена Галилео Галилея (1564—1642), а идея о том, что не только палец давит на стол, но и стол давит на палец была неподъёмна для учёной публики во времена Исаака Ньютона (1642—1727).

Объяснить мировоззренческие изменения 21 века сложно. А без этого непонятно, в чём обвинять просвещённое и вроде как рационально (по меркам прошлого века, но не нынешнего) мыслящее население глобуса. Ведь ещё пару десятков лет назад они и были носителями state-of-the-art мышления! Они были лучшими! Но state-of-the-art, «лучшее на данный момент» жёстко привязано ко времени. То, что хорошо вчера, уже не так хорошо сегодня: в трансдисциплинах всё время появляются новые приёмы мышления, новые объяснения реальности.

Новое содержание старых мыслительных трансдисциплин интеллект-стека сегодняшним образованным людям непонятно, как непонятна была ньютоновская физика во времена Нью-

⁸⁶ https://ru.wikipedia.org/wiki/Теория_резонанса

⁸⁷ https://ru.wikipedia.org/wiki/Теория_молекулярных_орбиталей

тона – и нельзя было даже объяснить, зачем она была нужна, аристотелевской физики ведь вполне хватало! Да ещё и Галилей со товарищи внёс в физику много нового (в том числе сделал физику экспериментальной наукой), прогресс физики был налицо, зачем вся эта новомодная ньютоновщина-лейбницеvщина в те далёкие времена? Все эти новомодные «интегралы» – зачем?! Мысль о том, что этому будут через сотню-другую лет учить в средней школе всех подряд, учёным того времени даже в голову не могла прийти.

Сегодня появилось новое знание, и даже не по физике и не по математике, а по мировоззрению, лежащему в их основе – и ему нужно опять всех учить, и учить всех подряд, в школе, вузе и за их пределами. Промывать населению глобуса мозги, вымывать из них мировоззренческие флогистоны и витализмы прошлого двадцатого века. Нужно ли продолжать учить физике и математике? Да, но и в них тоже другому содержанию: они тоже поменялись, и существенно.

Трансдисциплины: название то же, содержание уже другое

Как соотносимся мы, любимые, физический мир вокруг нас и его модели/теории (ментальные, компьютерные и даже физические)? Слово «солипсизм» тут не надо вспоминать, волнует более насущное: вот у двух разных инженеров разные информационные модели одной и той же атомной станции – как их соотнести друг с другом? А у двух менеджеров разные версии вроде бы как одной и той же методологии управления проектами разработки софта SCRUM – как им договориться? Как мы узнаём, какие модели/теории мира верны, и можно ли вообще говорить о «верности» моделей? Опять же, речь не идёт о «научных доказательствах». Всё много прозаичней: постановка диагнозов, предсказание погоды, моделирование беспилотных автомобилей, предпринимательские гипотезы. И конфликты, возникающие у людей по их поводу, и бесконечные переговоры в попытках совместить разные варианты моделирования одной и той же ситуации. Интеллект-стек в его современном виде – это прикладное мировоззрение, ни разу не «история философии»! Так, слово «онтология» в нём ровно то же, что в 1920 году, но содержание дисциплины полностью другое!

Кто бы мог подумать, что работы Е.Т.Jaynes от 1998 года формально (то есть математически! доказательства с формулами!) покажут возможность вероятностного вывода/обновления по теореме Байеса из книжки 1763 года, известной в современной формулировке, сделанной Лапласом в 1812 году! И это даст мощный толчок работам в области искусственного интеллекта, и выдвижению гипотезы о том, что человеческий мозг – это именно байесовский вычислитель? Кто бы мог подумать, что работы, в которых показаны ограничения суждений о мире на основании чисто статистики (а хоть и «новомодной» байесовской, а не намертво устаревшей традиционной) будут выполнены уже в 21 веке. Ключевая работа Judea Pearl с доказательством того, что причины и следствия нельзя вывести из данных, и для рассуждений о них должны рассматриваться контрфактические рассуждения, вышла в 2000 году, пересмотрена в 2009 году и пересказана простым языком для широкой публики только в 2018 году как «Книга Почему: новая наука причины и следствия»⁸⁸.

Эти идеи «причинной революции», изменившие статистику (можно после них выкидывать большинство статистических расчётов в здравоохранении, например, включая расчёты по обоснованию финансирования крупных эпидемиологических мероприятий типа локдаунов и поголовного вакцинирования), работают не только для искусственного интеллекта, но и для естественного тоже!

Эти новые идеи описывают логику науки с её гипотезами и свидетельствами/экспериментами много точнее, чем все предыдущие теории науки. Заявления типа «экстраординарные заявления требуют экстраординарных доказательств» получили в том числе и своё количественное, хотя и вероятностное обоснование. Увы, хотя этому знанию аж двадцать лет, просвещённое ещё в 20 веке учёное население глобуса настолько привыкло к вековой поступи прогресса в области логики, что до сих пор не замечает эти драматические двадцатилетние изменения! 1763-1812-1998-2009-2018 – всё это были частные исследования особо яйцеголовых, «широко известных в узких кругах». Проблема в том, что это лучшее на сегодняшний день знание о том, как мыслить о причинах и следствиях – и этому знанию хотя бы в начальной форме нужно учить всех.

Теория вероятностей сегодня – это абсолютно другая теория вероятностей, чем была ещё в 2000 году! В исследованиях по машинному интеллекту байесовская теория вероятностей, байесовская наука – это уже мейнстрим, иначе бы тамошние проекты были бы безуспешны,

⁸⁸ Judea Pearl, Dana Mackenzie, The Book of Why: The New Science of Cause and Effect, <https://www.amazon.com/Book-Why-Science-Cause-Effect-ebook/dp/B075DCKP7V/>

просто не работали бы. А что фронтир? Фронтир говорит, что должна быть **квантовоподобная/quantum-like вероятность**, а не байесовская. У квантовоподобной вероятности есть как минимум два достоинства по сравнению с байесовской: она считается быстрее (поэтому биологические системы используют именно её), а ещё она не отбрасывает то, по чему априорные значения неизвестны, а как-то это учитывает⁸⁹.

А вузовские учебные программы, если не брать программы по машинному интеллекту? Учебные программы пока не откорректированы. Но вот **вы прочли сейчас об этих изменениях, вы предупреждены, вы можете принять меры по собственному образованию, образованию друзей и сотрудников.**

Сообщество машинного интеллекта (именно инженеров AI интересуют вопросы мировоззренческого уровня – им нужно разобраться в них настолько, чтобы научить и развить не просто уже заведомо умных людей, а заведомо тупые компьютеры) продолжает поставлять новые знания на замену прежних «незыблемых истин». Так, успели отменить принцип «корреляция двух переменных не позволяет судить о том, изменение какой переменной является причиной изменений другой переменной». Объяснения (они ж всегда про причинность!) и доказательство влияния чего-то на что-то теперь никогда не будут прежними – но об этом пока мало кто знает. The Book of Why как раз об этом. Вы будете её срочно читать и перечитывать, или так и будете жить с допотопным пониманием причинности из 20 века? Знание причинности кардинально изменилось в мире где-то в 2009 году (научные работы, например, Judea Pearls), а «попсовый пересказ» в виде The Book of Why вышел только в 2019 году. Так что, будете читать?

Такие же интересные истории можно рассказывать про 4D экстенционализм, который позволяет более-менее компактно записывать происходящие во времени изменения, быстро договариваться о процессах и их представлении, согласовывать понимания разных людей в моделях деятельности. Более-менее строгим рассуждениям о системах в физическом мире учат буквально в паре мест в мире, число выпускников – пара сотен человек на много миллиардов жителей Земли. Книжка Криса Партриджа BORO book⁹⁰, рассказывающая об этом и рекомендуемая нами, практически неизвестна широкой публике. А ведь без неё очень трудно понимать системный подход, и трудно увязывать знания разных дисциплин о физическом мире в одно целое. При этом фронтир тут тоже поменялся, и теперь речь идёт о конструктивных онтологиях, которые части-целые представляют не как объекты в отношении часть-целое, а как конструируемые какими-то операциями объекты – и это даёт возможность думать и о частях-целых для воображаемых/математических/ментальных объектов⁹¹.

О «бизнес-процессах» трудно договориться, если не знать идеи 4D экстенционализма, изменения ведь трудно «увидеть», видят-то люди вещи и привыкли обсуждать «вещи», процессы трудны для мышления. Книга Криса Партриджа ведь как раз об этом. Идея проста: все изменения в мире происходят от взаимодействия физических объектов-индивидов. Следовательно «процесс» определяется входящими в него вещами, изменяющимися во времени. Приходите на производство и проверяйте: входят ли окружающие люди и вещи (оборудование, материалы) в процесс – и вы мгновенно договоритесь о границах процесса, о его особенностях. Если не будете привязывать процесс к физическому миру, а наоборот, будете пытаться использовать «классификаторы» (то есть повышать, а не понижать уровень абстракции) – не договоритесь никогда, что мы постоянно и видим в жизни. При этом может оказаться, что даже

⁸⁹ <https://pubmed.ncbi.nlm.nih.gov/33347968/> и <https://www.frontiersin.org/articles/10.3389/fnbot.2022.910161/full#B45>

⁹⁰ <https://yadi.sk/i/2SgjuILB3PqJEZ>

⁹¹ <https://ailev.livejournal.com/1653296.html>.

несколько сот лет существующую систему двойной итальянской записи в бухгалтерии придётся переделывать, она тоже уже не современна⁹²!

То, что системный подход помогает бороться со сложностью – это вроде как все признали. Но он сегодня по факту неизвестен: системное мышление представляется широкой публике или «систематичным» (то есть скрупулёзной проработкой всех клеточек в какой-то неведомой таблице, обходом всех веток каких-то неведомых деревьев решений), или «холистичным» с невнятными призывами тщательно думать, поскольку «всё со всем связано». Это и так понятно, что нужно тщательно думать, но как это конкретно помогает бороться со сложностью?! Ответа на этот вопрос нет, если не заглядывать в учебник современного системного мышления. Но ведь в этот учебник мало кто заглядывал! Этой дисциплине на русском языке ни в школе, ни вузе не учат! Впрочем, и на английском языке этому учат только системных инженеров, но со сложностью в проектах должны ведь совладать все! Тут нельзя думать, что можно сдвинуть всё это системное мышление на AI: без системного мышления вы ведь просто не догадаетесь задать хороший вопрос, поэтому не получите хорошего ответа. Как когда-то шутили про теорию решения изобретательских задач, «ТРИЗ помогает хорошему инженеру, а плохому инженеру не помогает», так и тут: AI как очень умный собеседник помогает другому умному собеседнику, а дураку общение с очень умным собеседником не поможет, он не догадается поговорить о важном, или просто не поймёт, что там важного в ответах, которые ему говорит умный собеседник.

Системный мыслитель про весь мир думает как про наборы вложенных друг в друга и взаимодействующих на каждом уровне матрёшек системных уровней – да ещё и много матрёшек в каждой, да ещё и эти матрёшки непрерывно эволюционируют. Свойства этих матрёшек на каждом уровне вложенности не совпадают со свойствами той матрёшки, в которую они вложены. И каждый уровень такой «мировой матрёшки» (верхняя матрёшка – вселенная с галактическими кластерами, нижняя матрёшка – какие-нибудь квантовые суперструны, но большинство интересного происходит на небольшом числе средних уровней физических объектов – живые клетки, организмы, космические корабли, транспортная инфраструктура Земли представляют собой очень узкий диапазон размеров) обслуживается своим экспертным сообществом, разобравшемся, как этот уровень матрёшки устроен. Понимание, как этот уровень матрёшки устроен, меняется каждые несколько лет, эволюцию знания не остановишь. Вот если у тебя есть такой «матрёшечный» взгляд на мир, и ты ещё привык сначала смотреть на ту обычно находящуюся в чужих руках матрёшку, в которую вкладывается находящаяся у тебя в руках матрёшка, и только потом раскрывать свою матрёшку-в-руках – вот тогда ты системный мыслитель. Всё остальное – бантики и рюшечки, мелкие уточнения.

Уже в самом конце 20 века при разбирательстве с этими матрёшками-системными-уровнями появилось понятие «проектная роль» (матрёшки же всегда у кого-то в руках, и они нужны им для чего-то!), и это отразило идущий в философии уже сто лет «прагматический поворот». И уже в 21 веке этот ход на прагматизм и проектные роли был довершен: жизненный цикл стал восприниматься не как набор работ, а как набор практик – деятельности людей над частями этих матрёшек, и сразу стало понятней с мультидисциплинарностью в проектах, а само системное мышление после обсуждения уже в 21 веке начало рассматриваться как часть мышления сильного интеллекта, вошло в самые разные транисциплины интеллект-стека.

В 21 веке системный подход опять поменялся, в 2020 году он не такой, каким был в 2000 году, а в 2000 году он не такой, каким был в 1965, в 2030 году (немного уже осталось!) он будет не такой, как сегодня.

В интернете вы увидите огромное число учебников системного мышления, написанных по идеям 1965 года (первое поколение системного мышления), а не 2020 года (второе поколе-

⁹² <https://www.academia.edu/95095498>

ние системного мышления). Но самые интересные работы физиков и биологов по системному мышлению вышли в 2021—2022 годах, и системное мышление стало мышлением третьего поколения⁹³!

Но это полбеда. Беда в том, что призывы к изучению системного мышления есть, а массового обучения системному мышлению нет! Как будто это просто «мышление опытных людей», а не какое-то мышление, которому можно просто взять – и научить, как учат той же ньютоновской физике, или даже как в учебных программах по машинному интеллекту начинают потихоньку учить байесовской логике.

Алгоритмика была перепутана с компьютерной грамотностью, и бесконечно отстала от жизни. Основные вычисления населения идут сейчас в дата-центрах и иницируются через смартфон, но школьники продолжают учить, что «компьютер состоит из системного блока, клавиатуры, мыши и экрана». Где, где мышь в ноутбуках?! Какой алгоритм позволяет распознавать речь, мы ведь уже общаемся со смартфоном прямо голосом – без клавиатуры! По-прежнему этот гибрид пульта по вызову такси, телевизора, видеокамеры и персонального телеграфа называем «телефон», хотя звонит он уже крайне редко – общение уходит в чаты. Но как он работает? Для большинства людей на планете это чистая магия, и нет учебного предмета, который учит тому, что там происходит. Как взрывается атомная бомба – этому учим, а как переводчик Гугла переводит с фарси на испанский, и каждый год всё более профессионально, уж не хуже изучающего иностранный язык даже уже не школьника, а студента вуза – вот этому не учим, это абсолютная магия.

Алгоритмика живёт сегодня даже в сознании программистов, обученных десятком лет назад, только в виде невнятных идей. И тут ещё приходят квантовые компьютеры, которые имеют дело не с информацией в битах, а со сверхинформацией в кубитах. Об этом не могут говорить даже большинство профессиональных программистов, а с точки зрения обычных людей квантовый компьютер и вовсе шайтан-машина. Но ведь квантовый компьютер намерен быть ровно той шайтан-машиной, которая резко усилит совокупный интеллект человечества. Мы будем учить школьников в школе ньютоновской физике в подробностях, или расскажем хотя бы «на пальцах» принцип работы квантового компьютера? Вообще, кто и где будет давать понятие вычисления в общем виде в учебных программах? И программирование: как говорит Karpathy, самый модный язык программирования сегодня – английский, а программировать надо нейронные сети, настраивать на конкретную задачу универсальные алгоритмы (и это программирование собственной нейросети, программирование нейросети AI, программирование нейросетей ваших друзей и сотрудников и даже программирование коллективной нейросети вашей организации, которая состоит из нейросетей сотрудников и нейросетей в её софте). Современному человеку неплохо бы хоть как-то быть знакомым с происходящим, чтобы не оказаться в мире магии, где надо знать точное заклинание, чтобы вызвать демона из системы искусственного интеллекта, и дальше демон продемонстрирует чудеса, недоступные для понимания.

Если мы хотим, чтобы население как-то справлялось со сложностью окружающего мира, то нам нужно обучать это население (от школьников до профессоров) современным мыслительным трансдисциплинам. Обучать явно, быстро и эффективно. Мыслительные трансдисциплины ни разу не сухая теория, они практичны, они организуют продуктивную деятельность, они поддержат коллективное мышление.

Увы, тут нужно написать, что польза современных трансдисциплин «потенциальна», а не «актуальна». Актуально польза трансдисциплин проявится только тогда, когда понятия, собранность, семантика, математика, физика, онтология, алгоритмика, логика и так дальше по всему интеллект-стеку в его современной версии, овладеют массами. Не массы овла-

⁹³ <https://ailev.livejournal.com/1656653.html>

деют сильным мышлением, а сильное мышление массами, тут должно быть без иллюзий – массы должны быть завоёваны этим мышлением, сами массы добровольно овладевать новым мышлением не будут. Ждать этого овладения можно долго. Чтобы ускорить процесс, массы системному мышлению нужно учить. Ага, «жалеть и учить, жалеть и учить».

Можно говорить о том, что миру надо дать коллективный иммунитет от глупости⁹⁴. Ибо каждая новая версия интеллект-стека (он меняется ежегодно! Прогресс неостановим!) означает, что владеющие этой версией люди и машины (AI) умней, чем владеющие предыдущими версиями интеллект-стека. То есть эти агенты умнее, в буквальном смысле этого слова.

Так что вам придётся напрячь расслабленный поп-культурой мозг: это привет из вашего завтра, где конкуренция не даст скидки на неграмотность. Миром начинают править удивительно крепко подкованные в «технарском» мышлении люди. Они не ходят в костюмах, они ходят в футболках, а гуманитарии у них на службе и на их содержании.

Основная причина, почему скоропостижное обучение на «трёхдневных курсах» непрерывно появляющимся новейшим прикладным практикам типа TameFlow в менеджменте или Jobs-to-be-done в инженерии не помогает – отсутствие фундаментального образования в части трансдисциплин, то есть отсутствие мастерства рассуждать правильно (логика!), координировать свои действия с другими людьми (методология!), управлять вниманием (собранность! со времён изобретения йоги и даосских медитаций в этой предметной области тоже немало произошло, в том числе за последние двадцать лет), строить модели (семантика, онтология, рациональность!) и вычислять по этим моделям (алгоритмика!), давать оценки применимости этих моделей (исследования!), точно выбирать тот кусок мира, для которого строится модель и разворачивается вся деятельность (системная инженерия!), занимать свою роль в разделении труда (методология!), вовремя замечать неадекватность своих действий и исправлять ошибки мышления (логика!), не вредить при этом (этика!) ... этот список можно продолжать и продолжать, и этому нужно учить со школы, в явном виде.

Пока же мы имеем даже в лучших учебных заведениях образование от педагогов-геронтократов. Геронтократов не по своему возрасту, а по возрасту тех замшелых идей, которые они выучили от своих преподавателей, а те от своих. Педагогика/андрагогика/хьютагогика (включая тех инженеров и менеджеров, которые вдруг начали преподавать) сегодня преподаёт не state-of-the-art трансдисциплин, педагогика преподаёт традицию, догму предыдущих веков – никогда не преподаётся предмет, по факту преподаётся история предмета. Вот как философия: философствовать никто не учит (да это и не нужно никому!). Зато учат истории философии, вставили это в образовательные программы в обязательном порядке!

То же самое безумное отставание обучения от жизни происходит с любым предметом. Пустите инженера преподавать инженерию: он, как генерал, который всегда готовится к прошедшей войне, будет учить студентов разбираться с допотопными конструкциями, будет «история инженерии». Это счастье, если где-то в этой истории доходят до науки, инженерии, искусства конца 20 века. Но с конца 20 века прошли уже десятки лет! «За время пути собачка могла подрасти»!

Мировоззрение как текущий интеллект-стек всё-таки более-менее стабильно, оно существенно меняется за пару десятков лет. Это прикладные практики меняются каждые 4—5 лет, в соответствии с циклом хайпа Гартнер⁹⁵, в соответствии с ритмом смены информационных технологий, в соответствии с ритмом инвестирования во множество стартапов по всему миру. Фундаментальное знание трансдисциплин живёт дольше, на него можно опереться.

⁹⁴ <https://ailev.livejournal.com/1520730.html>

⁹⁵ <https://ru.wikipedia.org/wiki/Gartner>

Но нельзя опираться на своё фундаментальное образование больше пары десятков лет! Если вы более-менее выучились мыслительному мастерству к 22 годам, то в 42 года обязательно нужно учиться снова, и содержание вашего образования будет другим! Если вы вот прямо сейчас пройдёте наш курс с очень кратким описанием пока ещё наисвежайшей версии трансдисциплин интеллект-стека, сделайте пометку в календаре – пройти его ещё раз через пяток лет! Хотя сейчас всё так быстро меняется, что и три года тут может быть более полезным предложением. Трансдисциплины меняются так же быстро, как и любые другие прикладные дисциплины, это нельзя игнорировать.

Современное мировоззрение основной массы населения земного шара насквозь мифологично. Старинная аристотелевская логика – допотопный миф, давно заменена математической логикой. Старое системное мышление, в котором не было ещё людей в их разнообразных проектных ролях – антикварный миф. Старая алгоритмика, в которой ещё не было квантовых компьютеров – древний миф. Новой рациональности нужно учить всё население, от школьников до маститых учёных – хотя учёных будет учить ещё сложнее, им очень трудно принять роль ученика! Макс Планк ещё замечал: «Не следует думать, что новые идеи побеждают путем острых дискуссий, в которых создатели нового переубеждают своих оппонентов. Старые идеи уступают новым таким образом, что носители старого умирают, а новое поколение воспитывается в новых идеях, воспринимая их как нечто само собой разумеющееся»⁹⁶.

Мрачно пошутим, что ситуация с этим только ухудшилась. Никакого застоя в науке нет, по сравнению со временами Макса Планка всё только ускорилося, но зато есть увеличение продолжительности жизни носителей старых научных идей и увеличение их бюджетов за счёт денег налогоплательщиков. А поскольку речь идёт не больше и не меньше, как о мировоззрении, о мыслительных трансдисциплинах, то тут холивары могут идти столетиями. Это вам не смена теории флогистона теорией кислородного горения⁹⁷ буквально за несколько десятков лет!

Так что мы в 20х годах 21 века попали во что-то типа интеллектуального средневековья, только со смартфонами и искусственным интеллектом. Попытки заглушить телеграм и твиттер в самых разных странах, выпустить единый набор учебников для школьников всей страны как раз из этого разряда: все всё понимают, но против современной инквизиции ничего поделать нельзя – рациональные аргументы не работают, когда сознание широких масс (включая законодателей!) насквозь мифологично.

Ошибки не в самой политике, а ошибки в мышлении, в незадействовании современной версии мыслительных дисциплин для выработки политических решений. **Рациональность, которая вроде как была обретена человечеством в целом, оказалась эфемерным достижением – её вроде как уже и нет, просвещённые на базе старинных версий трансдисциплин и непросвещённые вообще без задействования трансдисциплин интеллект-стека мнения оказываются одного примерно качества, дикарского.**

⁹⁶ <https://ru.citaty.net/avtory/maks-plank/>

⁹⁷ <https://ru.wikipedia.org/wiki/Флогистон>

2. Понятизация

Быстрое мышление через поиск ассоциаций

Понятизация – это практика как-то находить предметы похожими друг на друга и на этой основе выделять какие-то «фигуры из фона», выделять объекты наблюдения и действия (смотрю на гвоздь и забиваю его, смотрю на палящее солнце и загораживаюсь от него) из огромного числа потенциальных окружающих объектов. А ещё эти объекты надо как-то называть (необязательно словами, это иногда называют маркировка/labeling/означкование), и поэтому общее название роли **поэт**, у которого практика «разглядеть что-то в пестроте будней – и как-то обозвать».

Метафору понятизации и связанных с ней трудностей можно найти на вот этой картинке (в каждой шутке есть доля шутки):



Тут нужно уметь ещё отличать имена предметов (name) от самих предметов (thing), обозначаемых именами, то есть уметь понимать разницу не только между крестами и кольцами (предметами), но и между крестами металлическими и крестами католическими (понятиями).

По большей части речь идёт о врождённой части нейросетевого (в том числе человеческого, но есть гипотезы, что вообще вся вселенная в каком-то роде огромная нейросеть) интеллекта: умение группировать воспринимаемые в физическом мире объекты и воспринимаемые «умозрительно/умочувствительно» понятия как объекты по их похожести, равно как и понимать, что за звуками речи и символами на письме скрываются понятия.

Именно врождённая понятизация лежит в основе быстрого мышления (мышления-1, режим работы мозга S1) по Даниэлю Канеману⁹⁸.

⁹⁸ хорошо описано у него в книжке «Думай медленно... решай быстро» – <https://www.amazon.com/Thinking-medlenno-reshay-bystro-Russian/dp/5170800533/>



В машинном интеллекте это «быстрое» понятийное мышление хорошо реализуется глубокими нейронными сетями (deep neural networks). Но необученная нейронная сеть бесполезна в плане мышления примерно так же, как новорождённый ребёнок. Чтобы размышлять даже «быстро», «интуитивно» (про логичное мышление пока молчим), людям с их «мокрыми» нейронными сетями и искусственным нейронным сетям на самой разной аппаратной основе (классические компьютеры, оптические компьютеры, квантовые компьютеры, аналоговые мемристорные компьютеры и т.д.) нужен некоторый жизненный опыт, насмотренность/наслушанность/наработанность. Нужно долгое (для людей – многолетнее) и дорогое познание/исследование – предобучение, а потом ещё и настройка на ту или иную предметную область (профессионализация).

Маленький ребёнок насматривается, наслушивается, наигрывается – он тренирует нейронные сети в своём мозгу. Это неважно, что его не учили писать, читать, логично размышлять. Говорить ребёнок сможет, какие-то проблемы решать (обобщать примеры ситуаций, виденные им в жизни на новые ситуации, в чём-то похожие на старые) ребёнок тоже сможет. Даниэль Канеман подчёркивал, что режим S1 вполне обеспечивает речь, это вовсе не только «образное мышление», «визуальное мышление». Символические рассуждения в S1 вполне возможны! Невозможно только «алгеброй гармонию поверить», если что-то «показалось» или «почудилось», то в этом режиме нельзя логически покритиковать «привидевшееся» и отвергнуть. Хотя интуитивно и это можно, умение вести логические рассуждения появляется как эмерджентное свойство, оно тоже может познаваться/learn на примерах, и потом обобщаться.

S1 обеспечивает быстрые, неточные, с большим числом логических ошибок результаты. Кроме того, S1 не может объяснить полученные результаты. Но интуиция, «нюх» (включая самые разные математические, логические, физические интуиции), самые разнообразные ассоциации – это всё S1 поддерживает.

Если человек или компьютер с нейросеткой много видел, участвовал во многих ситуациях, много читал, много смотрел видео (и не только художественных фильмов, но и докумен-

тальных фильмов), то речь у него будет богатой, ассоциации точными и не ограниченными одной модальностью восприятия (видео, аудио, кинестетикой, вкусом, запахом).

Сколько это – много? Это зависит от размера нейросетки. Есть работа⁹⁹, показывающая зависимость размера нейросетки от оптимальной «насмотренности»: если насмотреться меньше, то аппаратные возможности недоиспользованы, если насмотреться больше – результата не будет, только зря потраченное время на познание, «некуда запоминать результаты». Если очень грубо, то для 40—70 млрд параметров нейросети для обучения надо предъявлять последовательности из 1.4 триллиона токенов (токен – это какой-то элемент как потенциальный носитель смысла, например пиксель, воксель, аудиоотсчёт, буква, слог или часть слова – корень, суффикс, приставка). Мощность человеческого мозга по отношению к оптимальности практически неограничена, нужно довольно долго (сейчас – порядка двадцати пяти лет) насматривать/нарабатывать/начитывать мозг, чтобы он начал выдавать приемлемые результаты мышления. Но нет насмотренности – привет на работе пятилетнему ребёнку, он тоже человек, просто насмотренности поменьше!

Большие языковые модели (large language models), с которыми работают сейчас в области AI, тут мало отличаются: они начитаны, удивительно творческие, хорошо пишут школьные сочинения и даже журнальные статьи, но удивительно глупы: в сгенерированных ими текстах есть ошибки, и эти сетки не трудятся их обнаружить и исправить. Со временем эта ситуация исправляется, ошибок у нейросетей меньше и меньше, и у взрослых людей ошибок меньше и меньше, но всё равно это не полная безошибочность строгого математического вычисления.

Первая же догадка, которая приходит «на ум» (живому человеку или компьютерной нежити) выдаётся как результат. Если догадка была в связи с очень похожей на уже встреченные в жизни ситуации (опыт имеет значение!), то всё ОК. Если ситуация отличается, то вероятность ошибки догадки тем больше, чем больше отличается новая ситуация от ранее встреченных. Так что начинаем мы с практики понятизации, как самой основы мышления, но практики абсолютно недостаточной для качественного мышления.

К слову сказать, и людей, и нейронные сетки учат сейчас распознавать и такие объекты как «ошибка»: люди могут буквально физически чувствовать при этом «ошибку» (например, чувство какого-то «дребезга» при сопоставлении слова и подразумеваемого им типа, например, чёткое ощущение чего-то не того в фразах типа «все три зверя чувствуют себя хорошо: слон, муха и арбуз». Идёт замедление хода мыслей, обрабатывается вот это ощущение «что-то тут не то!»).

#

⁹⁹ <https://arxiv.org/abs/2203.15556>, пояснения на русском с https://t.me/gonzo_ML/1216

Передача субъективного опыта

Модальности восприятия – это восприятия от наших пяти чувств: видео (зрение), аудио (слух), кинестетика (ощущения в теле – и тут много всего разного, включая проприорецепцию/проприоцепцию), ольфакторное (обоняние), густаторное (вкус). В какой модальности происходит понятизация? Ибо мышление идёт вроде как «в понятиях», амодально. Понятизация – это как раз «ввод-вывод для мышления», как раз про проявление абстрактных понятий в сознании, осознанности «содержания мышления».

Понятизация поэтому существенно связана с органами чувств, а также нашими представлениями «в голове» о нашем восприятии. «Внутри головы» мы визуализируем мысли в форме образов, аудируем в форме звуков, воображаем вкусы и запахи, а также представляем себе ощущения – «воображение работает ровно таким же образом, как и восприятие. Нельзя ничего почувствовать из того, что нельзя воспринять органами чувств». Но то, что мы можем воспринять извне, всё это мы можем и «вообразить», «галлюцинировать», «представить». Все эти «голоса внутри головы» и «картинки на внутреннем экране» реально существуют и представляют собой вполне работающие дополнительные интерфейсы к мозгу-вычислителю, как нейронной сети. Понятизация активно это использует.

Простейшая работа с нашим восприятием заставляет задуматься, как же мы представляем себе какие-то объекты. Например, в курсе системного фитнеса просят обратить внимание на кинестетическую модальность – представить себе ощущение обмякания каких-то мышц. Надо именно напрячь-расслабить какое-то место в теле, попробовать дорасслабить его вдогонку естественному расслаблению – и запомнить ощущение как «кинестетическую gifку, маленький фильм дорасслабления». А затем обращаться с этим воспоминанием как словом/знаком на «телесном языке» (всё есть текст!). Что же тут может пойти не так? Удивительно много:

- Можно запомнить дорасслабление как визуальный образ, не обращая внимания на ощущения тела. Дальше теряем время: вспоминаем визуальный образ, потом рефлекторно переходим от него к «мышечной памяти» и только тут начинается действие.
- Запоминаем ощущение в теле, но оно запоминается не как «развёрнутое во времени изменение в ощущении», а как «статичная фотка».
- Запоминаем не ощущение, а последовательность действий, приводящая к этому ощущению, и туда наверняка попадает что-то лишнее. Например, можно сразу расслабить-дорасслабить место в теле. Но запоминаем почему-то сначала напряжение, которое нам нужно было только для того, чтобы начать расслабление. Его не нужно, если уже место в теле расслаблено, а дальше надо только дорасслаблять сознательно. Но мы сначала вспомним последовательность операций, затем выполним лишнее напряжение, а потом уже начнём дорасслаблять.
- ... ещё много подобных ситуаций. Люди плохо работают с кинестетической модальностью: не обращают внимание на ощущения в теле. А ведь интуиция часто проявляется в кинестетической, а не аудиальной или визуальной модальностях, и уж совсем редко в форме запаха или вкуса!

Так что нужно тренировать работу с кинестетической модальностью – как восприятием мира в **первом доступе** (терминология взята в адаптации её в НЛП из теории автоматов¹⁰⁰). **Прямой доступ** – это то, что поступает на датчики, то есть фотоны в глаза, перепады давления воздуха в ухо и т.д., дальше это кодируется датчиком и перерабатывается

¹⁰⁰ <https://ailev.livejournal.com/649230.html>

мозгом. Человек получает первый доступ в момент осознания восприятия, то есть человек осознаёт только результат обработки прямого доступа. Когда речь идёт о коммуникации, даже о попытке описать свои собственные восприятия первого доступа, мы выражаем это восприятие привычным нам языком – словами и жестами, картинками. В попытках описать восприятия первого доступа (ко внешнему миру или воображаемому/ментальному «внутри головы», это без разницы) мы уходим довольно далеко от точности выражения этого восприятия, если будем использовать слова, обозначающие какие-то абстрактные объекты, например «я ощущаю затруднение» (какое именно? что трудно?). Но мы можем поднять точность выражения, если будем заботиться о сенсорной обусловленности описаний (то есть отслеживать, что мы представляем мир довольно точно таким образом, как он выглядит, слышится, ощущается, пахнет, каков он на вкус – и ещё это в развёртке динамики, «gifками во всех модальностях восприятия»). И тут оказывается, что наличие языка и какой-то «насмотренности» накладывает ещё один фильтр: **сенсорная обусловленность** в понятизации профильтровывается только тем, что есть в культуре, **культурной обусловленностью**, поскольку, если чего в культуре нет (например, нет слов для передачи какого-то ощущения), вы передать его не сможете.

Предположим, у вас нет слов для передачи вашего ощущения для чего-то очень специфического из вашей практики – например, ощущения от расслабления чего-то в районе горла. Это называется «субъективный опыт». Передача субъективного опыта – это огромная проблема для философов. Но поскольку мы тут не занимаемся философией, мы эту проблему будем решать инженерно, «из первых принципов», то есть исходить из физики (и поступать так же, как поступают физики).

Роджер Желязны описывал, что два мага с изумлением обнаружили, что они видят заклинания по-разному: для одного это были ниточки с узелочками, а для другого лучики с блёстками. Вот это оно и есть. При этом, конечно, для визуальности у нас полно средств описать, что там видно – пока вы не увидите что-то типа вертолёт, который нужно будет описать Чебурашке, как в том анекдоте: «как бы тебе это объяснить? Апельсин знаешь? Ага, знаешь. Ну так вот вертолёт на него абсолютно не похож». А теперь опишите соседу, как вы чувствуете тальк, рассыпанный по полу – это ж просто кинестетика, да? Тело, которое вдруг начинает вести себя на скользком тальке совсем не так, как на земле, но и не совсем так, как на льду. И вы это тело воспринимаете не снаружи глазами, а в ощущениях, изнутри. Как это ощущение описать? Ощущение поцелуя, ощущения эээ... ну вы понимаете.

Итак, вы называете нужное вам ощущение «расгорлить» и понимаете, что никакие слова передать это ваше субъективное ощущение (субъективный опыт) другому человеку не помогают. Чтобы «расгорлить» (то есть представить это ощущение там, где вам надо – и получить эффект расслабления там, где надо), надо это представить – но передать из мозга в мозг ощущение нельзя! А уж если надо передать ощущение «расгорлить» компьютерной нейросетке, то это и подавно оказывается невозможным.

Тут помогает та же процедура, что известна физикам, когда они передают свои абстрактные понятия типа «ускорение свободного падения». Они задают какие-то операции во внешнем мире с понятными хорошо определёнными предметами, чтобы точно воспроизвести какой-то эксперимент. В эксперименте предусмотрено измерение, то есть обращение внимания на какой-то конкретный параметр системы в ходе взаимодействия объектов этой системы. Эксперимент уточняется и уточняется, пока не получается однозначным (например, не просто кидается дробинка и пёрышко с Пизанской башни, а падение делается в трубке, из которой откачан воздух – уточняются условия, убираются мешающие факторы, добиваются однозначного воспроизведения у разных людей).

Как понять, что такое красный? Выполните инструкцию: проколите на солнечном свету (конечно, тут мы поступаем как физики: долго описываем все условия эксперимента, необходимые и достаточные, чтобы получить результат измерения) пальчик булавкой, посмот-

рите на это место, там будет жидкость как раз красного цвета, хотя и разных оттенков. Или купите в магазине красную краску (попросите продавца продать «красную краску») и поглядите на неё. Ещё лучше выполнить пять-шесть разных замеров разными способами, чтобы как-то обобщить результат. Ощущения, понятно, запоминаются – и маркируются словом. Но передаётся от источника знаний не «краснота», а инструкция по её получению. Когда мы обсуждаем, как в эволюцию/развитие включить развитие внутреннего опыта, если мы не можем записать это знание в мемом где-нибудь рядом в книжке, то это тупик. Мы должны в мемом включать маркер для названия ощущения («красный») и пару-тройку инструкций по выполнению практики, как это ощущение получить. Другой объект: запоминается практика со знакомыми объектами, чтобы получить незнакомый внутренний опыт. Хотите узнать, что ощущает гимнаст во время сальто? Сделайте сальто, и вы это ощутите!

Не заметили подвоха в предыдущих строчках? Чтобы сделать сальто, нужно что-то сделать со своим телом, для этого нужно как-то им проуправлять (как? вы ж не делали никогда сальто?), а уж потом вы попадаете в сальто – то есть вы как гимнаст воспроизводите телом ощущения прыжка в сальто (но оно вам неизвестно!), а потом уже вы можете ощутить, что там в самом прыжке, если вы ухитрились туда как-то допрыгнуть. То есть сначала опытный гимнаст должен передать вам соматомеханическое описание (что там происходит с телом изнутри тела: сома – это как раз тело, как оно чувствуется изнутри тела) для захода в биомеханическое внешне видимое сальто вашего тела, чтобы потом вы оценили, как там оно в этом прыжке в ощущениях. То есть вы уже должны неплохо владеть сомой, чтобы выжимать из сомы всё новые и новые ощущения. Ну да, вы уже должны знать про зрение, чтобы смотреть на «красный», а не слушать «красный», или не пробовать «красный» на вкус. Всегда есть какое-то базовое знание, вы обычно не с полного нуля начинаете, у вас уже есть какая-то насмотренность/наслышанность/начувствованность, а также наработанность как опыт манипулирования какими-то предметами, включая собственное тело.

Как тренеру системного фитнеса продемонстрировать обмякание в ваших перенапряжённых мышцах, чтобы дальше вы могли работать с этим ощущением? У всех эти мышцы перенапряжены в разных местах, внутри себя все чувствуют разное (описывают это как жжение, давление, «тянет», «прёт», набухает, холодит и т. д. – и часто это ещё и совсем другие ощущения). Всё просто: вы проводите ряд операций типа подавить рукой на стенку с разной силой, а потом прекратить давление – и обратить внимание на ощущения, да ещё сознательно попытаться его усилить, «доотпустить мышцу», сопроводить этот сброс усилия. Не всё, конечно, вот прямо так просто (то есть как и в случае «выйди на солнышко перед тем, как смотреть на пальчик», чтобы было точно «красное», а не «в темноте все кошки серы», нужно выполнить несколько дополнительных условий, повышающих вероятность того, что вы ощущаете ровно то, что нужно), и давите затем вы ногой, лбом – пытаюсь обобщить это ощущение сброса усилия мышцами. Запоминаете это под названием «обмякание». Пытаетесь воспроизводить с самыми разными регионами тела (вам же даже названия мышц не нужны!), поднимаете вашу чувствительность к «обмяканию» (новое ощущение! но вы уже знаете, каково оно, поэтому можно захватить его вниманием и отслеживать его маленькие изменения!), добиваетесь беглости воспроизведения сброса усилий мышц в разных частях тела «по памяти» (то есть не надо давить куда-нибудь, чтобы получить ощущение обмякания, а нужно просто дать сознательную команду – «сбрасывай усилие», мышцы обмякнут, и вот оно – ощущение «обмякания», воспроизведённое по вашей уникальной памяти). Как записать это уникальное знание «обмякания», чтобы потом обсуждать его, передавать от человека к человеку, оно ж абсолютно уникально ощущается в каждом мозгу, в каждом теле?! А вот так, процедурой «подготовка измерения, проведение измерения, передача результата измерения в память». И, конечно, «результат измерения по заданной процедуре» получает название, которое можно использовать при обсуждении. Вот это самое «расгорлить» можно передать по вот этой процедуре.

Возьмём случай посложнее. Вот собранный человек с хорошо развитым сознанием, которое понятийно наводит внимание и умеет каким-то образом не терять это внимание подолгу. Те, кто занимаются просветлением, берут коан (любая бессмысленная мысль, смысла которой понять заведомо невозможно – вам же нужно занять ум задачей, у которой заведомо нет решения, чтобы задача никогда не кончилась! Смысла заведомо не должно быть, чтобы вы вечно могли его искать!) и удерживают его во внимании. Оказывается, это можно тренировать, как цирковой трюк: удерживать размышление над бессмыслицей весь день, а потом ещё и полночи (быстрый сон), а потом еще и всю ночь (медленный сон). Ура, после круглосуточного удержания внимания ваш мозг попадает в особый режим круглосуточной работы сознания, вы можете называться «пробуждённым» (другое название того же самого – «просветлённый»), ибо вы спите, но ваше сознание как удержание внимания на задаче работает. Полностью бессмысленное дело, цирковой номер, ни на что не влияет, занимает кучу времени, но вам же хотелось узнать «каково оно, быть пробуждённым/просветлённым» изнутри? Вот, как раз операционный рецепт, цифровая (точно воспроизводимая, словесная, знаковая) память, нужная для эволюции, простота репликации, точность повторения (это повторяли сотни тысяч людей! Просветлённых на Земле и сейчас десятки тысяч в странах, где практикуют буддизм, индуизм, ламаизм). Нужно только выполнить последовательность операций, потратить от двух до шести лет жизни. Эффектов для интеллекта от этого особо никаких не получите (скорее, наоборот – число лауреатов нобелевской премии в странах, где много просветлённых ниже, и вообще уровень жизни ниже, если бы люди от такого умнели, всё было бы по-другому¹⁰¹), хотя внутри мозга это очень, очень приятно.

Альтернатива: удерживать внимание путём записей (на бумаге или в компьютере – это уже становится неприципиальным), записи помогут удержать внимание на каком-то предмете и на пять лет, не только на время сна! А за приятными ощущениями проще сходить в баньку попариться, или послушать хорошую музыку, покушать клубнику со сливками и т. д. Можно понять попов/жрецов и прочих продавцов религиозных мемов: чтобы затащить к себе, они дают подобные ощущения «кайфа от нестандартных режимов работы мозга», и человек на них подсаживается – всё, вот ещё один монах, который искренне рассказывает, что можно получать кайф вот таким извращённым способом. Репликация мемов практики работы с вниманием! Но можно ведь без кайфа и цирковых эффектов удержания мозга в несколько лет тренируемых режимах, а просто быть осознанным и собранным для работы. Более того, можно этими альтернативными «техническими» методами удерживать собранность и коллектива! И даже не надо заботиться о телепатии, если полагаться на смартфон, который по большому счёту эту проблему «передачи мысли на расстояние» решил, и решил надёжно, дёшево и без затрат на обучение и тренировки. Этот инженерный подход в сто раз быстрее и в тысячу раз полезнее. Поэтому субъективно ощущаемая «собранность» передаётся тоже словесно, тоже записывается – но эта собранность (включая собранность, поддержанную экзкокортесом/моделером) как субъективное ощущение не может быть описана, как она ощущается изнутри. И всё же она описывается: как набор операций, приводящих вас в состояние, когда вы просто её ощущаете, то есть становитесь собранным.

Для «расгорлить» вы придумываете последовательность упражнений, в которых внимание того, кому вы хотите это передать, направляется на какое-то расслабление («подними гирию правой рукой на упоре, затем отпусти гирию – почувствуй, что у тебя расслабляется», «повтори другой рукой» и т.д., далее инструкция на генерализацию – всё это обобщается как «расслабление где-то»). Дальше можно дать какие-то упражнения, где мышцы напрягаются уже не гирей – внимание знает, как искать где-то в теле «расслабление». Далее надо задать упражнение, которое чуть-чуть напрягает мышцы в горле, которые надо расслабить. И потом сознательно надо

¹⁰¹ <https://ailev.livejournal.com/428032.html>

усилить это расслабление «вдогонку», то есть «дораслабить» сознательно. Полученное конечное ощущение после действия сознательного «дораслабления» в горле – это и есть результат операции «расгорлить». Всё, достаточно теперь сказать, что надо «расгорлить», и мы получим искомое состояние, просто сознательно выполнив это «дораслабление в нужном месте» без предварительных нагрузок и других упражнений. Время изготовления навыка какого-то телесного действия, время использования навыка – получения нужного состояния, достижения заданных ощущений, индивидуально воспринимаемых каждым в их первом доступе, но трудных в словесном выражении.

Выполнить всё это оказывается возможным, если мы даём инструкции на сенсорно-обусловленном, а также культурно-обусловленном языке (не обязательно словами! Кинестетика, если эти ощущения нам знакомы, это тоже текст, причём «слова» там могут быть «гифками», то есть это всё динамическое, а не статическое. Слова речи «в голове», они ведь тоже «аудиогифки», меняющиеся во времени фонемы, а не «один статичный постоянный звук»!).

Конечно, тренировать сенсорную обусловленность и преодоление культурной обусловленности нужно для всех модальностей, не только кинестетической. Например, хорошо бы понимать, что происходит в визуальной модальности – «картинки внутри головы» не самый плохой интерфейс к нейросети, эти картинки вполне могут становиться сами по себе объектами внимания точно так же, как и внешне воспринимаемые зрительные образы. Но как и с кинестетическими «образами», нельзя считать, что мышление идёт именно в визуальной модальности. В европейской культуре визуализации «внутри головы» обычны для мыслящих людей, с детства они работают с диаграммами и художественными сложными образами, схемами и иллюстрациями. В культуре индейцев визуализация мышления – это дар богов, или нужно пить психотропные вещества, чтобы тебя «посетили видения». В книге «Визуальное мышление. Доклад о том, почему им нельзя обольщаться»¹⁰² рассказано, что полагаться именно на визуализацию, как основной способ мышления, неправильно:

¹⁰² https://ridero.ru/books/vizualnoe_myshlenie/



https://ridero.ru/books/vizualnoe_myshlenie/

#

Нейросемиотика. Нейросемиотическое программирование

Понятийное мышление вполне синестезийно в части представления его результатов как доступного восприятию, то есть которые можно выделить вниманием, но абстрактно/внемодально в части понятийной работы, не привязано к каким-то определённым модальностям. Понятия представляются как некоторые области многомерного пространства, отвечающего всем возможным понятиям. И эти понятия могут затем отображаться как картины (или даже «фильмы/гифки»), звуки (или даже «речь»), ощущения (или «кинестетические гифки»), и т.д., или даже «всё это вместе и одновременно, тесно переплетённое», то есть синестезийно¹⁰³ и идеастезийно¹⁰⁴.

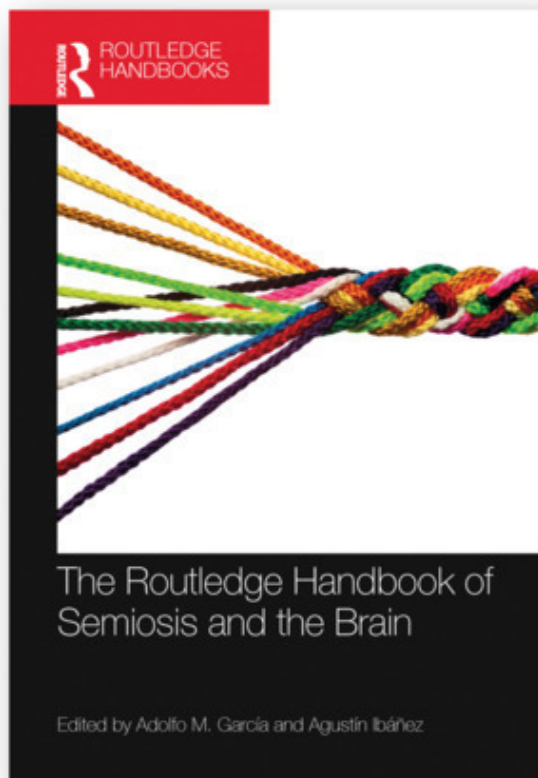
В машинном интеллекте те же тренды: языковые модели получают уже и для визуальных наборов данных, и наборов данных из подписанных картинок. Всё больше и больше исследователей замечают, что есть огромное сходство старого нейролингвистического программирования (NLP, neuro-linguistic programming)¹⁰⁵ и современной работы с нейронными сетками в части естественного языка (NLP, natural language processing, сводимого сегодня к «пониманию человеческой речи на естественных языках»). Но так же, как понятийное мышление не визуально (ну, или «и визуально тоже», можно визуализировать результаты), оно не аудиально, то есть необязательно идёт «словами». Поэтому один из подходов тут может считаться нейросемиотическим/neurosemiotic¹⁰⁶ – это про то, каким образом формируются понятия и знаки в нейронных сетях, в том числе человеческом мозге, но и не только – компьютерные нейронные сети сюда тоже попадают, а ещё можно рассматривать и другие варианты вычислителей (семиотика – это наука знаках).

¹⁰³ <https://ru.wikipedia.org/wiki/Синестезия>

¹⁰⁴ <https://ru.wikipedia.org/wiki/Идеастезия>

¹⁰⁵ <https://libcat.ru/knigi/religioznaya-literatura/samosovershenstvovanie/390938-dzhon-grinder-shepot-na-vetru.html> – описание подхода на русском языке, от одного из соавторов подхода, профессора лингвистики.

¹⁰⁶ The Routledge Handbook of Semiosis and the Brain, 2022: Neurosemiotics aims to identify and explain the neural processes and mechanisms that manifest as sign-relations in the animal kingdom, thereby supplying semiotic models to neurobiological accounts.



Модель мира отражается в многомерном ментальном пространстве понятий, а потом как-то становится доступной сознанию (иногда причудливым образом: уши слышат, как наш рот проговаривает какие-то слова – одна часть мозга генерирует речь, а вторая часть мозга обеспечивает восприятие этой речи как бы извне, но вот сама понятийная работа, само мышление как поиск решений проблем остаются «невоспринимаемыми» – я специально не употребляю тут модально-окрашенных слов «невидимыми», «неслышимыми», «не унюхиваемыми»).

Понятизация занимается тем, как соотносятся объект, его знак, понятие, значение, смысл в нейросети живой или не очень живой, или даже другом вычислительном субстрате **с позиции восприятия этого вычислителя** («я работаю с понятиями»), как вообще идёт означкование в каком-то вычислителе, какие алгоритмы можно применить в этом вычислителе для работы с понятиями. **Нейросемиотика** – это про понятизацию в нейровычислителях (в том числе человеческом мозге, но это могут быть и сообщества, и общества). **Семантика** занимается соотношениями между знаками, стоящими за ними понятиями и объектами, которые эти понятия отражают, причём с внешней позиции восприятия субстрата вычислителя («он работает с понятиями», а не «я работаю с понятиями»), вычислитель и его природа тут не так важны, сколько важно, с чем этот вычислитель должен работать – с отношениями знаков, понятий, объектов. Как в обычном программировании/информатике можно выделить алгоритмику (качественные алгоритмы, выдающие правильный результат за минимальное время) и семантику как вопрос о смысле вычислений, так и в нейролингвистическом/словесном или нейросемиотическом/знаковом в любой форме программировании можно тоже выделить понятизацию/алгоритмику (что делать, чтобы быстро и правильно думать о понятиях, знаках и предметах и как-то устанавливать между ними соотношениями) и семантику (каковы должны быть соотношения между знаками, понятиями и объектами, чтобы они имели смысл). Программирование/обучение/преднастройка на контекст может быть как живого человеческого мозга (трудно представить мозг кошки, хорошо работающий со знаками), так это

может быть программирование/обучение/преднастройка на контекст нейронной сети в «неживых» вычислителях, «разговаривающий компьютер» уже вполне существует.

В какой-то мере самые разные последовательности не только букв, но и фонем в речи, нот в музыке, паттернов в чём угодно можно (и, наверное, нужно) считать текстом, следуя известной максиме Jacques Derrida¹⁰⁷ «всё есть текст». Если вам пару веков назад бросили перчатку, то это такой «динамический знак» (жест), означающий вызов на дуэль. Это «словарное значение» знака, но их может быть и несколько разных («косил косою косою косою»). Вокруг **текста** как последовательности знаков (или даже одного знака) есть **контекст**, который уточняет значение текста (или даже одного знака) – ту область пространства понятий, куда мы попадаем. А то, зачем мы вообще занимаемся вычислениями с этим текстом, задаёт **смысл**: какое отношение к изменениям в физическом мире имеют эти вычисления.

Пространство понятий в знаках выражается квантованно/дискретизовано, ибо часто значение находится где-то там, где нет подходящего знака для его выражения («между точками дискретизации пространства понятий»), и приходится работать с пространством понятий через грубый язык, через знаки, а иногда и изобретать знаки – быть поэтом. Вот это и есть предмет понятизации. Представьте себе 100 оттенков синего цвета, и слова-знаки «синий» и «голубой» (при этом в некоторых языках даже может не быть слова «голубой»). Вот это типичная ситуация: требуется довольно много слов, чтобы выразить некоторые понятия. Потом их можно означковать, но это если они часто встречаются. Большинство понятий не имеют каких-то своих значков, но вполне выразимы на естественном языке.

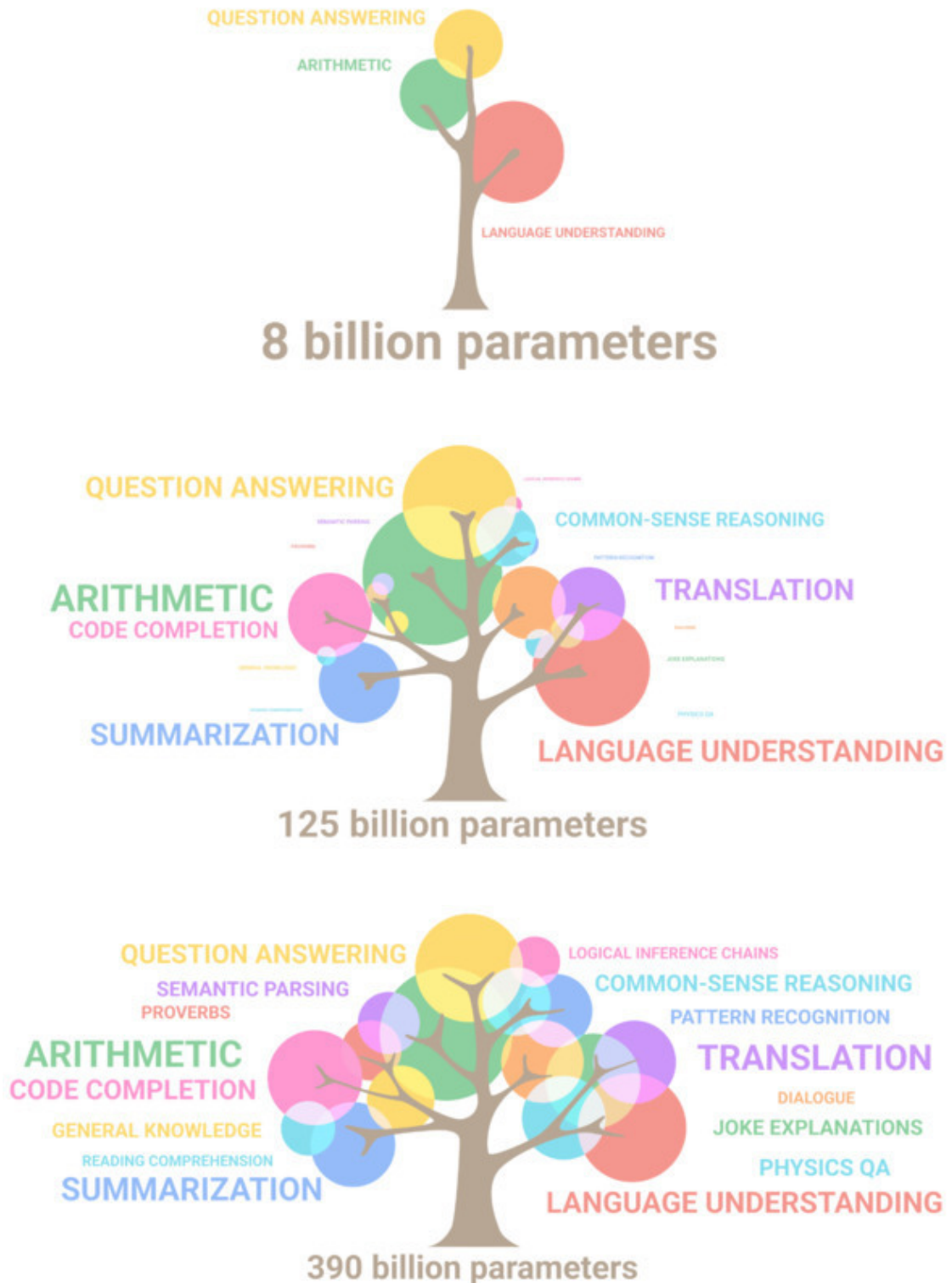
Нейролингвистические программисты влед за Хомским говорят о **поверхностной структуре** языка (выразимой знаками, как изображение пикселями, а звук отсчётами амплитуды в какие-то дискреты во времени) и **глубокой структуре** (понятия, кодируемые нейросеткой, эти понятия обычно «промеж пикселей, результат вычислений», не точно соответствуют каким-то знакам). Знаки (поверхностная структура) тем самым представляют какие-то вехи, обозначающие места в понятийном пространстве (глубокая структура), поэтому иногда говорят о знаковой координатной сетке/grid (скажем, понятия разных синих цветов вам или больше доступны, если есть слова/знаки «голубой» и «синий», или менее доступны – вы хорошо можете различать голубой и синий цвет, но не сможете это хорошо выразить. Чем больше развит язык, тем точнее можно высказываться на нём. Ньютон писал свои трактаты на латыни, ибо тогдашний английский был как язык довольно убог).

Представьте себе, что у вас есть мир Майнкрафта и слова для каждого его объекта. А потом вам предлагают описать ваше реальное рабочее место только этими словами. Трудно, да? В этот момент вам захочется и как-то аллегорически или метафорически заговорить на этом языке, а иногда и новые слова в этот язык внести, а иногда переопределить значение уже имеющихся слов. По большому счёту, с естественным языком происходит ровно вот это: выражение на языке знаков (без разницы, синестетических, визуальных, аудиальных, динамических/жестов, кинестетических и т.д.) – это дискретизация, что-то типа АЦП, аналого-цифрового преобразования, а понимание выраженного знаками – обратный процесс, «цифро-аналоговое преобразование» со всеми вытекающими особенностями аналогового представления (например, невозможность многократного точного копирования, ибо при аналоговом копировании накапливается ошибка).

По большому счёту, значительная часть работы человеческого и машинного (Bing, ChatGPT, Bard, ERNIE Bot и множество других систем AI на базе современных нейронных сетей) мыслительного мастерства сегодня проходит в простом и лёгком режиме S1 по Канеману. Из экспериментов в AI известно, что способности нейронной сети к обучению растут

¹⁰⁷ https://ru.wikipedia.org/wiki/Деррида,_Жак

с размером этой сети, и при достаточных размерах даже наблюдается рост способности к обобщению, а потом и возможность рассуждать¹⁰⁸:



¹⁰⁸ <https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>



Человек в этом плане лучше, чем муравей или кошка. В мозгу человека множество разных нейронных сетей в разных частях мозга причудливо связаны друг с другом так, что появляется и понимание языка, и понимание шуток. А в мозгу кошки нейроны связаны друг с другом так, что понимание языка и шуток не появляются. А вот в искусственных нейронных сетях это всё появляется, но эти сети тоже должны быть не любыми, а специально устроенными.

Сами по себе проверки на ошибки в каких-то представлениях в нейронной сети не появляются, и огромные нейронные сети страшно ошибаются, если решаемые ими задачи вдруг становятся проблемами, то есть требуют разбирательства с новой предметной областью. Ребёнок, у которого не удерживается внимание, который не владеет логикой, но который много чего повидал и почитал – вот это и есть человек с хорошо развитым мастерством понятизации и отсутствием разных других видов мастерства. Это поэт, который легко отождествляет морскую звезду из монографии по биологии морских беспозвоночных и Патрика из мультфильма Спанчбоб. Поэтому в его рассуждении морская звезда со дна моря вполне может сказать несколько слов и иметь друзей. Это поэзия, склеивание миров. В мире поэзии граф Дракула – вампир, а высказывание «вампиры не существуют» заставляет графа Дракулу смеяться. Поэт интуитивно может расклеить разные миры, но может и не расклеить – ибо устранение ошибок смешивания рассуждений из разных контекстов делается обычно логическими проверками.

Мы начинаем учиться тогда, когда уже умеем выделить какую-то фигуру из фона: выделить объект (ментальный или во внешнем восприятии) и обнаружить сходство его с другим каким-то объектом, который вынимаем из нашей памяти. Начинаем с врождённого умения выделять объекты и обобщать способ, которым мы эти объекты выделяем (находим в объектах похожесть). Это и есть понятизация – работа с понятиями, их ассоциирование/сопоставление друг с другом на предмет нахождения похожестей.

О том, что мышление человека существенно опирается на аналогии, на нахождение паттернов/закономерностей/связей/шаблонов/ритмов хорошо написано в книге Дугласа Хофштадтера «Гёдель, Эшер, Бах. Эта бесконечная гирлянда»¹⁰⁹. Это книжка аж 1979 года (написана сорок два года назад), поэтому она давно уже неактуальна как «передовое знание человечества», но она являет собой хороший сборник примеров причудливости проявления аналогий и ассоциаций, хорошее введение в саму проблематику понятизации. За сорок с лишним лет после издания этой книги она перестала быть «передовой», и человечество существенно продвинулось в объяснении того, что в этой книге описано. Но главное – у человечества появилось машинное обучение, а в рамках машинного обучения появились большие языковые модели (large language models), которые показывают, каким образом появляется

¹⁰⁹ https://ru.wikipedia.org/wiki/Гёдель,_Эшер,_Бах

понятизация, дают возможность экспериментировать и возможности инженерной разработки устройств, которые выполняют понятизацию.

Всевозможные детские упражнения на «логику» типа «какой тут предмет лишний» – это, по большому счёту, упражнения не на логику, а на классификацию (отнесение предмета к какому-то типу на основании его схожести по каким-то критериям). Это как раз про понятизацию. Логика нужна была бы только в том случае, если нужно было бы объяснить, почему было принято то или иное решение по поводу отнесения объекта к типу, но для этого объяснения нужно будет много чего ещё уметь, кроме как выдать догадку, что же в ряду лишнее или что на что похоже.

Классические тесты на IQ – это тесты на часть понятизации, связанные с определением паттернов. Это хорошо показано в работе François Chollet «On the Measure of Intelligence»¹¹⁰ (2019). По большей части это тесты, которые не требуют даже владения естественным языком¹¹¹ (хотя там для людей бывают и лингвистические тесты, но это не главное там тестируемое, и компьютерные варианты этих тестов не обращаются к знанию языка для тестируемых на интеллект алгоритмов¹¹²).

IQ тесты мало тестируют мастерство в других практиках интеллект-стека, хотя они и связаны тесно с собранностью (если у вас внимание отсутствует, то вы просто не помните, какой паттерн уже видели – и ничего повторяющегося в мире у вас нет. То есть для понятизации уже нужна собранность! Деление на отдельные практики интеллект-стека и выстраивание в последовательность довольно условно). И хоть понятизация лежит в основе интеллект-стека, это не самая большая его часть. Так что IQ не слишком коррелирует с успешностью в науке или бизнесе, хотя некоторая связь, конечно, есть. Но не такая прямая, как это обычно представляется. Люди со средним IQ могут получить хорошее образование и добиться больших успехов в науке. А люди с высоким IQ вполне могут иметь мастерство в других практиках интеллект-стека похуже, и не достичь особых результатов. Интеллект не определяется полностью качеством именно понятизации или собранности. Интеллект определяется совокупным мастерством мышления по дисциплинам всех практик интеллект-стека в целом.

#

¹¹⁰ <https://arxiv.org/abs/1911.01547>

¹¹¹ <https://lab42.global/arc/core-knowledge/>

¹¹² <https://lab42.global/arcathon/>

Психопрактики для понятизации: внимание к кинестетике

Помощь в понятизации людям может прийти со стороны психопрактик фокусирования/focusing (вытаскивание инсайтов из телесных/кинестетических ощущений)¹¹³ и вытекающих из них практик концептуализации этих инсайтов thinking at the edge (ТАЕ)¹¹⁴. По своим идеям эти практики очень близки к нейролингвистическому программированию, но выходят за классические рамки НЛП в том, что не работают чисто с «процессом», оставляя «содержание» где-то в глубинах нейросетки. Нет, это практика понятизации как работы с содержанием: «процесс» как раз вытаскивает содержание из нейросетки, вытаскивает субъективный опыт (первый доступ «внутри головы») – и пытается передать его другим людям, найти выражение для этого опыта.

ТАЕ как практика начинает с внимания к ощущениям тела (это тоже знаки! Всё есть текст!), затем эти ощущения стабилизируются, затем они выводятся на уровень понятий (осознание того, что эти ощущения могут означать), затем этим понятиям даются имена.

ТАЕ заходит довольно далеко и в практики интеллект-стека, которые находятся после понятизации, ибо кроме концептуализации/формализации/моделирования на поздних этапах там включается и логика тоже. Focusing и ТАЕ продолжают линию метафора-ориентированной психотерапии, выводя **терапию с пациентами (ремонт поломанных людей)** за пределы терапии, превращая её в **инженерию с клиентами (не поломанными!), обучение с учениками (не поломанными!)**. Авторы ТАЕ пишут, что все «шаги» их процедур нужны только для того, чтобы научить. А потом «немые обретают язык, и выражают свою интуицию быстро и понятно». Это ровно то, что нам нужно на этапе осознания: из синестезии по линии либо визуальной (образы, всякие техники «ответа на экране»), либо кинестетической (как в психопрактике фокусирования, предтече ТАЕ) получить какие-то тексты, характеризующие содержание пространства смыслов – для той его области, в которой часто и слов-то нет для обозначения новых, пока безымянных концептов. В ТАЕ оговаривается, что даже при отсутствии слов для концептов можно использовать всю полноту естественного языка для выражения этих новых концептов, так что это тоже не препятствие.

Аналогичные ТАЕ практики легко представить и в других модальностях (визуальной, аудиальной), не только кинестетической, как в оригинальном методе. В принципе, такое можно делать и в синестезии, это ведь тоже тренируется. Безмодальное/внемодальное пространство смыслов отражается в сознании синестезийно, и потом только осознаётся разными модальностями – ощущениями в теле, мелькающими перед внутренним взором картинками, слышимыми внутренним слухом звуками и т. д. Вот и использовать всё это богатство внутреннего представления мира. Модели мира внутри себя сознанием ощущаются ведь ровно так же, как мир снаружи: видео, аудио, ощущения и т. д. – только это «внутри головы». И даже пространство «внутри головы» такое же, четырёхмерное (3D и время).

Классическое нейролингвистическое программирование¹¹⁵ жёстко критикуется психотерапевтами (почти всё содержание статьи в Википедии – это критика со стороны психотерапии), и часть этой критики верна. В том числе верна и в том, что это – не терапия, а больше инженерия. То, что предлагало NLP для людей, очень напоминает prompt engineering¹¹⁶ для современных нейросетей: настройка/дообучение вычислений нейросети в мозгу (S1) с понятиями в контексте, задаваемом «подсказками»/промтами/prompts, причём без особого соот-

¹¹³ <https://www.b17.ru/article/46839/>

¹¹⁴ <http://previous.focusing.org/tae-intro.html>

¹¹⁵ https://en.wikipedia.org/wiki/Neuro-linguistic_programming

¹¹⁶ https://en.wikipedia.org/wiki/Prompt_engineering

несения с S2 (особо на это «программируем непосредственно S1, прямо обращаемся к вычислителю для S1, не трогаем S2» были направлены паттерны так называемого «нового кода»¹¹⁷ NLP). Многие сотни известных приёмов и практики (паттернов) нейролингвистического программирования¹¹⁸ вполне получают объяснение в свете современных воззрений на работу нейронных сетей, хотя как и с любым знанием, за почти полвека с момента появления тамошних практик их набор надо существенно чистить (напомним, это инженерия, а не «чистая наука о мозге», но инженерные практики меняются едва ли не быстрее, чем научные теории, к нейролингвистическому программированию это тоже относится в полной мере). Причём это «программирование»: в NLP это называется «работа по процессу, а не по содержанию» (внимание уделяется тому, «как думать», алгоритму, а не «о чём думаем» – ровно как в компьютерных программах на традиционных языках программирования внимание уделяется алгоритму, а не содержимому переменных, оно будет определяться потом, в ходе работы, когда входные данные определяют значения переменных программы). Как ни странно, это не самая простая идея: в мире программирования разделение алгоритма (в NLP – процесса) и данных (в NLP – содержания), работа с алгоритмом как данными и данными как алгоритмами появилось не сразу и эти идеи могут принимать множество самых различных форм, учитывая и размытость между софтом и аппаратурой. Это верно и для людей: кошку и ребёнка можно научить каким-то условным рефлексам, но людей потом можно учить, просто сообщая им какие-то фразы, например, «не влезай – убьёт» (но для этого надо сначала людей научить языку, а кошку нельзя научить языку в силу плохой аппаратуры её мозга).

Оригинальное НЛП из 80х годов содержит некоторое количество инженерных решений в части моделирования «языковой модели внутри человека», а затем и в части обучения человека практикам программирования такой модели (обучения мозга, про плохую привычку: «как ваш мозг сумел это выучить?» – это не терапевтический, а инженерный вопрос!). НЛП было озабочено пониманием того, как работают результаты словесного программирования нейронных сетей людей, нейро (мозг) лингвистического (словами) программирования. А в современных искусственных нейросетках даже нет такого аспекта «работы с человеком, но не пациентом». Так что современная понятизация как-то должна объединить находки нетерапевтической/инженерной работы по моделированию человеческого совершенства (human excellence, предмет НЛП) и моделированию компьютерного совершенства (computer excellence, предмет computer science, hardware and software engineering и отдельных дисциплин типа artificial intelligence, deep learning и т. д. – всё, что работает с нейро-вычислителями).

¹¹⁷ <https://www.newcodenlp.org/>

¹¹⁸ Shlomo Vaknin, «The Big Book of NLP, Expanded: 350+ Techniques, Patterns & Strategies of Neuro Linguistic Programming», 2010, 828 страниц

3. Собранность

Агенты

Всё больше и больше физиков, которые вдруг занялись физичностью жизни, включая физичность интеллектуальной жизни, прибегают к использованию физики частиц¹¹⁹. Частицы/particles – это с одной стороны **части**, с другой стороны – «что-то особенное, обособленное от остального». В физике такие объекты обычно называют **системами/systems**: они состоят из чего-то внутри себя (подсистем, частей систем) и находятся в окружении/среде (environment), которое тоже состоит из систем – и имеют границу, которая различает их и окружение (то есть не расплываются в среде, а как-то поддерживают свою «самость», отдельность, автономность).

Физики описывают состояния этих систем, давая эти состояния как точки в пространстве состояний (не только точка положения частицы в 3D пространстве, но и температура, цвет, даже структура их строения из подсистем, а также какие-то характеристики производных по времени – скорость, ускорение и т.д., всё это может быть размерностями пространства состояний). И вот дальше частицы могут быть инертными по отношению к изменению своих состояний и активными – и некоторые частицы настолько «странны», что планируют изменения своих состояний, ставя целью достижение выгодных для них состояний и при этом намеренно проходя состояния, которые для них менее выгодны, чем исходные – ожидая в конечном итоге получить результат лучше, чем у них сейчас. Грубо говоря, эти частицы имеют межвременные предпочтения и готовы чем-то пожертвовать сейчас, чтобы получить результат потом. По факту это поведение деятеля – он планирует путь через пространство состояний, предполагающий достижение конечной потенциально выгодной точки, хотя на этом пути могут встречаться и какие-то отдельные ухудшения. Такие «странные» автономные частицы, способные к планированию, называют **агентами**¹²⁰.

По сути дела, в таком подходе победили панпсихисты¹²¹ в той их ветви, где живое и неживое, а потом сознательное и несознательное представляют собой некоторый континуум, а не противоположности¹²². То есть в природе существует «нечто», которое как-то ухитряется поддерживать свою отдельность по отношению к окружению – хоть атом, который довольно долго живёт не распадаясь, хоть капля масла в воде, хоть инфузория туфелька, хоть кошка, хоть человек в батискафе на дне Марианской впадины. Частица, или чувствующее (sentient) существо, или даже разумное существо с сознанием.

Примерно та же линия рассуждения об эволюции у Виталия Ванчурина¹²³ со товарищи, но там чуть другие слова: про многомасштабную/multiscale эволюцию и IPU/information processing unit как «единицы информационной обработки». Астрофизики (они и сами вполне себе физические объекты, IPU) всполошились, что системная (в том числе инопланетная) жизнь проходит мимо них, и выдали статью про планетарный интеллект¹²⁴. Это продолже-

¹¹⁹ <https://arxiv.org/abs/2210.12761> от Friston со товарищи, <https://arxiv.org/abs/2301.10077> от Ванчурина со товарищи, <https://www.mdpi.com/1099-4300/25/1/129> вот от от Hoffman со товарищи, <https://writings.stephenwolfram.com/2023/02/computational-foundations-for-the-second-law-of-thermodynamics/> от Wolfram, constructor theory <https://www.constructortheory.org/> по факту про то же.

¹²⁰ <https://arxiv.org/abs/2210.12761>

¹²¹ <https://plato.stanford.edu/entries/panpsychism/>

¹²² <https://arxiv.org/abs/2112.15242>

¹²³ https://scholar.google.com/citations?hl=en&user=nEEFLp0AAAAJ&view_op=list_works&sortby=pubdate

¹²⁴ <https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/intelligence-as-a-planetary-scale->

ние линии рассуждений Вернадского и гипотезы Гайи¹²⁵ (при всей её критике), работ Varela и Maturana по autopoiesis¹²⁶, идеи complex adaptive systems (CAS) /complexity science¹²⁷. Статья приходит к странным выводам о том, что в атмосфере цивилизация проявляется в запрещении выпуска фреонов и прочему влиянию на атмосферу (для астрофизики это важно: состав атмосферы можно измерить!). To conclude, an exploration of an exploration of planetary intelligence can draw together three domains of study: the evolution and function of Earth's biosphere; the current emergence of the technosphere in the Anthropocene; and the astrobiology of worlds inhabited by technologically capable exo-civilizations.

Такого сорта статьи о Земле-существовании физически вполне осмыслены, хотя вполне могут быть поставлены на одну полку с текстом о специальном компьютере Deep Thought, который в результате семи с половиной миллионов лет непрерывных вычислений наконец-то выдал ответ на ultimate question of life, the universe, and everything: «42». Deep Thought предложил создать другой, ещё более великий компьютер, который будет включать в себя живых существ как часть вычислительной системы, чтобы узнать, в чём, собственно, состоит Вопрос. Этот компьютер был назван Земля и был настолько огромен, что некоторыми по ошибке воспринимался как планета. Сами исследователи, которые управляли программами, выглядели как обычные мыши. Так что астрофизики в своём тексте просто начали догадываться о том, что давно было уже известно даже мышам¹²⁸!

Вообще, в этом направлении «агенты в физике» типичны статьи типа написанных Artemy Kolchinsky¹²⁹ «Semantic information, autonomous agency and non-equilibrium statistical physics» и «A thermodynamic threshold for Darwinian evolution». Все эти статьи говорят об одном: понятие агента вполне физично, оно не требует какой-то теории витализма¹³⁰, признания особости живых разумных существ как агентов в части описания их поведения как чего-то, не укладывающегося в законы физики, существование каких-то трансцендентных (по ту сторону физического мира) свойств агентов.

Есть множество самых разных определений для агентов, но все они сводятся к тому, что речь идёт частицах, которые ставят целью улучшить своё состояние, активно влияя как на себя, так и на окружающий мир – и для этого они имеют модели себя, модели окружающего мира и занимаются постановкой целей и планированием достижения этих целей. И, конечно, сразу заходит речь о том, что агенты могут быть разной степени разумности. И вслед за этим приходится обсуждать самые разные смежные с **агентностью** системы понятия. Вот глоссарий технологического подхода к вездесущности разума TAME (Technological Approach to Mind Everywhere)¹³¹, там всё крутится вокруг понятия «Я»/Самость/Self, как некоторой автономной сущности (частицы) с различными свойствами, и в самом глоссарном определении видно, что это «Я» имеет тип «система», то есть что-то более-менее устойчивое, отделяемое от окружающей среды, и далее мы будем постепенно разбираться с этими понятиями в самых разных практиках интеллект-стека:

- **Агентность/agency** – набор свойств системы, тесно связанных с принятием решений и действиями по приспособлению к обстоятельствам. Агентность определяет степень, в которой оптимальные способы взаимодействия с этой системой (в терминах коммуникации, пред-

process/5077C784D7FAC55F96072F7A7772C5E5

¹²⁵ https://en.wikipedia.org/wiki/Gaia_hypothesis

¹²⁶ <https://en.wikipedia.org/wiki/Autopoiesis>

¹²⁷ https://en.wikipedia.org/wiki/Complex_adaptive_system

¹²⁸ https://en.wikipedia.org/wiki/List_of_The_Hitchhiker%27s_Guide_to_the_Galaxy_characters#Deep_Thought

¹²⁹ https://scholar.google.ru/citations?hl=en&user=RmRwJJIAAAAJ&view_op=list_works&sortby=pubdate

¹³⁰ <https://ru.wikipedia.org/wiki/Витализм>

¹³¹ <https://www.frontiersin.org/articles/10.3389/fnsys.2022.768201/full>

сказания и контроля параметров состояния) требуют моделей все более высокого уровня, определенных в терминах масштаба целей, стрессов, возможностей и предпочтений этой системы как воплощенного в ней «Я», действующего в различных проблемных пространствах. Это представление об агентности связано с представлениями о том, что некоторые системы могут перестраивать себя (autopoiesis, «создание себя»)¹³² и они могут предвосхищать изменения в себе и окружении (anticipatory systems)¹³³.

- **Сознание/consciousness** – феноменальный опыт от первого лица любого «Я» – то, что делает мою зубную боль несводимо отличной для меня от зубной боли любого другого человека или описания зубной боли от третьего лица. Степень и содержание сознания – это «каково это» быть этим «Я», в отличие от изучения его извне, независимо от того, достаточно ли развито это «Я», чтобы быть в состоянии что-то об этом сказать или даже как-то об этом думать. Сознание здесь не обязательно означает развитое, рефлексивное, вербальное самосознание, а скорее как чувство из цикла «чувство-обработка/вычисление-ответ», которое рассматривается как что-то непрерывное от самых примитивных проявлений у клеток или даже компьютерного AI до полноценного сознания, каким оно известно у человека. А поскольку все мыслящие/cognitive агенты неизбежно состоят из частей, то и любой человек тем самым будет коллективным интеллектом – то, каково быть «Я-человеком», в точности соответствует тому, каково быть «Я-хитро организованный коллектив не очень осознанных клеток».

- **Мышление/cognition** – вся деятельность (включая действия не только «ума», но и «тела» – в кавычках тут только потому, что это необязательно мышление человека!), осуществляемая «Я» любого масштаба и варианта материального воплощения, которая лежит в основе сбора, обработки и действия с информацией в целях приспособлений к условиям окружения и по возможности превентивной защиты от рассеивания (буквально: рассеивания молекул, смерти) системы этого «Я» в окружающей среде. Мышление включает активное (то есть с возможным включением операций с телом и окружающими предметами, меняющее ситуацию) умозаключение, познание/learning/обучение и базовую целенаправленную деятельность по изменению себя, мира, моделей себя и мира, а также сложные мыслительные навыки, такие как символические рассуждения, составление концепций, язык и мета-познание (познание того, что такое познание).

- **Решение/decision** – событие во время прохождения состояния системы через некоторое релевантное пространство (помним, что это не только 3D пространство, но обобщённое пространство каких-то состояний, каждое из которых может быть описано многомерным вектором в пространстве состояний), которое эффективно моделируется как выбор между различными вариантами целевого состояния. Степень «принятия решения» любой данной системы пропорциональна пространственно-временному (в 4D) и сложному (с учётом других характеристик пространства состояний, не только пространства-времени) расстоянию между событиями, которые в конечном итоге приведут к определенному результату, и самим результатом. Продвинутое «Я» учитывают в их механизме принятия решений контрфактические будущие состояния. Речь идёт прежде всего об опоре в решениях на объяснения.

- **Разум/ум/mind** – функциональный, динамический аспект «Я», являющийся результатом всех его мыслительных и соматических (сома – тело, как оно ощущается «Я» изнутри этого тела) действий, который представляет собой склонность к определенным типам действий в перспективе от первого лица, и эта разумность сохраняется при изменениях тела. То есть это определённого сорта «вычислитель» (хотя в эти вычисления может входить и тело!).

- **Интеллект/intelligence** – функциональная способность «Я» решать проблемы в различных пространствах состояний (не обязательно в трехмерном пространстве), не привязан-

¹³² <https://en.wikipedia.org/wiki/Autopoiesis>

¹³³ [https://en.wikipedia.org/wiki/Anticipation_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Anticipation_(artificial_intelligence))

ная к конкретным реализациям «Я», анатомическим структурам или временным шкалам. Сила интеллекта пропорциональна умелости в ориентировании и перемещениях по этим пространствам, **включая, в частности, способность определять пути, которые временно ведут дальше от состояния цели, но в конечном итоге позволяют достичь лучших результатов**. Сила интеллекта оказывается несводима к результатам простых классических IQ-тестов на обнаружение паттернов и удержание внимания на паттернах, она отражает не только наличие адекватной «аппаратуры» (фактор G) для вычислений интеллекта, но и результаты познания агентом окружающего мира и познания агентом себя. Продвинутой интеллект использует дополнительные уровни само моделирования, что позволяет осуществлять многоуровневое само моделирование «Я» и внешнего мира в его возможных будущих состояниях (контрфактическое мышление), а также включает тревожность и креативность (выявление возможных проблем, в отличие от решения только существующих сейчас проблем). Интеллект тоже «безмасштабный», ибо не существует разумного агента/системы, который не состоит из частей, поэтому интеллект всегда коллективный, и нет отдельного вида «Я» во многих системных уровнях, который только и будет обладать интеллектом.

- **Иерархия потребностей Маслоу/Maslow's hierarchy of needs** – мотивационная теория в психологии, которая фокусируется на относительных типах предпочтений и целей, которые преследуют человеческие (или другие) системы на различных стадиях и масштабах наблюдения (Maslow, 1943)¹³⁴. Она также подчеркивает степень интеграции подсистем и модуляцию высших уровней уровнем стресса для частей. Помним, что мы пересказываем тут глоссарий ТАМЕ, поэтому сохранили тамошнюю отсылку к довольно старой идее иерархии потребностей. Вполне возможно, что в более свежих версиях ТАМЕ эта отсылка будет заменена на какую-то более современную.

- **«Я»/Самость/Self** – целостная система, состоящая из совместно действующих (интегрированных в неё) частей, которая служит функциональным владельцем ассоциаций, воспоминаний (то есть это не эргодическая система из физики, а имеющая память) и предпочтений (имеет предпочтения находиться в каких-то точках пространства состояний), которая действует для достижения целей в конкретных проблемных пространствах, где эти цели принадлежат коллективу подсистем как целому «Я», а не какой-то отдельной подсистеме. «Я»/самость определяется пространственно-временным масштабом (системным уровнем) и характером типов целей, которые она может преследовать (то есть каких мест в пространстве состояний она может достичь, и по каким путям) – ее «мыслительным световым конусом»¹³⁵. «Я» имеют функциональные границы и материальные реализации, но не идентичны какому-либо конкретному типу несущей «Я» системы и могут пересекаться с другими «Я» на том же, более высоком и более низком системном уровне (то есть уровнях «часть-целое»). «Я»/самость – это теоретическая конструкция, представляемая внешними по отношению к «Я» системами (такими как ученые, инженеры и другие «Я») и самими системами по отношению к себе (через внутренние модели «Я»). Концепция «Я» облегчает прогнозирование изменений и приспособление к изменениям в окружении и самой системе-носителе «Я», сохранение «Я» служит эффективной высокоуровневой целью для стратегий вмешательства и контроля в «естественный» (без участия «Я») ход событий.

- **Стресс/Stress/неудовлетворённость** – удалённость целой системы «Я» от желаемого нахождения в каком-то желаемом/целевом месте пространства состояний, запускающая активные (мыслительные, включая телесные) действия по попаданию в это желаемое место пространства состояний, чем больше стресс, тем более активные действия предполагаются (хотя при огромном стрессе состояние такое, что действовать уже невозможно: система ломается,

¹³⁴ https://en.wikipedia.org/wiki/Maslow%27s_hierarchy_of_needs

¹³⁵ https://en.wikipedia.org/wiki/Light_cone

так что лучше бы ей действовать заранее, пока стресс не так велик). Пространственно-временной масштаб и масштаб сложности событий, которые могут вызвать стресс в системе, являются хорошим показателем мыслительной сложности системы. Стресс может возникнуть из-за несоответствия между внешними состояниями и потребностями «Я», между воспринимаемым миром и собой и ожиданиями того, что должно было бы восприниматься, или между целями нескольких подсистем внутри агента, как на одном, так и на разных уровнях организации. Таким образом, неустроенности/geometrical frustrations¹³⁶ между системными уровнями, представления материаловедов о нагрузке (и возможной из-за неё поломке) как высокоуровневым факторе, влияющем на поведение системы во времени – это минимальные примеры фундаментальной концепции стресса. Стресс в его общем виде определяется так же, как метаболический стресс у бактерий, конкурирующие силы выравнивания клеток при образовании тканей и «настоящий психологический стресс» у разумных организмов.

Повторимся, что все эти важные понятия в разных школах мысли определяются немного (а иногда и совсем) по-разному, у каждого из них есть множество синонимов, но общая суть остаётся:

- Есть некоторый набор частей, которые в каком-то смысле «разумны» (но необязательно это осознают, хотя могут выполнять довольно сложное поведение, причём обучаться этому поведению), и из этих частей каким-то образом создаётся система, у которой есть «Я». То есть любой интеллект оказывается коллективным, собранным из частей.

- Эта система «Я» по мере роста разумности не просто избегает плохих для себя состояний «меня съели» или «я умер от голода или холода» (по большому счёту, это и есть «цель жизни» – «выжить сейчас и в будущем»), но и начинает предсказывать наступление этих состояний (тревога и креативность) и планировать, а также выполнять действия, которые уводят от этих состояний, возможно, проходя состояния ещё хуже текущего (агентность).

- Часто в таких рассуждениях указывается, что клетка печени лучше выживает, если не мешает организму в целом добиваться улучшения условий существования организма – тогда её шансы уцелеть (и шансы её генов быть отреплицированными в печени новых поколений организмов) сильно растут.

- Это всё абсолютно физично, эти все рассуждения не зависят от масштаба (клетки, организмы, популяции) или вида субстрата (живые существа, роботы), на котором реализованы «Я». Но обязательно, чтобы были датчики состояния среды и себя, эффекторы, влияющие на среду и себя, возможности вычислений, а также память (без которой обучение невозможно). Более того, это не зависит и от масштаба времени (филогенез как «развитие вида систем» и онтогенез как «развитие организма из генома или технической системы из мемона» рассматриваются одинаковым образом, разве что «время развития системы» и «время эволюции системы» по их масштабу могут различаться в тысячи раз).

¹³⁶ https://en.wikipedia.org/wiki/Geometrical_frustration

Собранность: владеть умом и телом

Итак, мышлением как деятельностью, посвящённой достижению цели занимается агент, который:

- Сам по себе внутри себя имеет разных других агентов «поглупее», которые играют в нём какие-то роли, заняты выполнением тех или иных функций и имеют какие-то свои цели – и цели верхнеуровневые им могут быть в том числе и неизвестными (или известными, и тогда требуют принятия решений: соглашаться или не соглашаться, ибо если «моя рука вдруг меня не слушается, делает что-то своё» – это может привести к изменению поведения верхнеуровневого агента).

- Планирует свой маршрут через пространство состояний и отслеживает его прохождение – и тут важно, что это не всегда приятное путешествие (меняется мир вокруг, меняется своё собственное состояние, а также меняются модели мира/окружения и себя, и меняются ожидания от прохождения ранее придуманного плана в новых обстоятельствах). Тут нужно ещё и указать, что часто приходится идти в прямо противоположном направлении, не теряя из виду общей цели.

- Состоит в составе «объемлющего агента», как-то соотнося свои цели с целями этого наагента (скажем, понимая, каких целей хочет добиться команда проекта, в которой участвует наш агент-человек).

Можно исходить из того, что агент, его подагенты и его наагент (все – коллективные интеллекты!) играют какие-то роли в одновременно идущих на разных уровнях не столько даже «спектаклях» (ибо там не только роли, но и реплики записаны), сколько в ролевых играх¹³⁷ (типа «дочки-матери», или ролевых играх по отыгрыванию каких-то миров, известных как «игры живого действия» где известны только роли и связанные с ними возможные паттерны действий, и только общая канва сюжета – и важно импровизировать все реплики, поведение и часто даже намерения роли).

¹³⁷ https://ru.wikipedia.org/wiki/Полевая_игра



Агент должен удерживать внимание на важных объектах всех этих многоуровневых отыгрышей ролей в разных играх (которые кажутся внешним наблюдателям «импровизационными спектаклями»). Агент не должен просто рассеяно блуждать по пространству состояний, ибо как в природе: зазевался – тебя съедят. Если вы пошли днём в магазин, чтобы не голодать вечером, нельзя забывать, что а) вы не просто так идёте, а чтобы вечером быть с едой и б) вы идёте в магазин, а не в цирк или просто дышите свежим воздухом: надо-таки не забыть дойти и купить еды. Если вы на большой скорости движетесь на автомобиле по хайвею, у вас нет автопилота, и ваше внимание блуждает – то от аварии вас отделяют считанные секунды. Всё это верно и для агентов-команд, агентов-предприятий: они не должны забывать, какими проектами заняты, какие цели преследуют. И не должны забывать, что это всё отыгрыш ролей – и роли можно менять. Если не удалось что-то сделать в одной роли, можно занять другую роль и попробовать другие паттерны действий. Если Василий Пупкин отыгрывает роль принца Гамлета, и вдруг вообразил себя и впрямь принцем Гамлетом, то он сумасшедший, а не «вошедший в роль». Это же верно для ролей инженера-электронщика, менеджера по развитию, бизнесмена: ежели Вася Пупкин вообразил, что он и есть инженер-электронщик, а не отыгрывает роль электронщика, то он сошёл с ума. Если кто-то вообразил, что отыгрывающий роль электронщика Вася Пупкин и есть электронщик – то да, на момент отыгрыша это правда, но когда Вася Пупкин пошёл обедать, то это не электронщик пошёл обедать, а агент-актёр. Это всё **«ролевое мастерство»**, сродни актёрскому мастерству как его учат в театральных училищах: играй, но не «перевоплощайся всерьёз», если тебе надо отыграть персонажа из сказки, или говорящее дерево, и ты забудешься в том, что это роль – жди беды. В жизни ровно то же самое со всеми другими ролями! Жизнь при таком описании становится многоуровневой ролевой игрой всех-со-всеми в самых разных ролях, отыгрыванием самых разных «миров».

Актёр/актор/агент, добивающийся успеха (регулярно избегающий попадания плохих состояний, потому как вовремя предпринимает действия по их избеганию) во всех своих многочисленных ролях во всех своих многочисленных проектах рабочих и личных (как играх живого действия, и этих проектов множество! должен быть **собранным**, то есть отлично владеть собой (и умом, и телом), он должен понимать, что происходит в проекте как игре

живого действия, кто ещё с ним играет в каких ролях, предугадывать их возможные действия (predictive coding¹³⁸ на многих уровнях), то есть иметь какие-то порождающие модели мира и себя, включающие достаточно деталей для точных предсказаний.

Понятизация учит обращать внимание на какие-то сложные ситуации в окружающем мире, называть объекты внимания в этих ситуациях, даже если по поводу этих объектов есть поначалу просто смутное ощущение. Собранность (и роль там «**собранный**», а поскольку задействуем экзокортекс, то иногда говорят «**киборг**») учит удерживать затем внимание на этих объектах, причём в ситуациях, когда «аппаратной поддержки» самого агента катастрофически не хватает для удержания этого внимания: попросту, собранность учит людей пользоваться для удержания внимания и себя, и коллективов ресурсами не только биологического мозга, но и экзокортексами в виде бумаги-ручки и бумажных книг раньше, компьютеров сейчас. Как в компьютере есть небольшая оперативная память и довольно обширная внешняя (дисковая) память и ещё большая глобальная (базы данных в датацентрах где-то интернете), так и в случае агентов-людей есть иерархия памяти:

- «мокрая» нейросетка головного мозга. Можно поднять вероятность того, что мозг что-то вспомнит вовремя, удерживая размышления на какую-то тему достаточное время. Например, можно что-то писать, тщательно формулируя (например, час править какой-то абзац содержательного текста). Это означает, что содержание понятого за тридцать секунд обдумывания абзаца вспомнится менее надёжно, чем содержание обдумываемого час абзаца. Как удерживать внимание на этом содержании на час? Записать, переписать, перевести на другой язык, разметить типы – любые операции, которые позволят удержать «мокрую» нейронную сетку человеческого мозга на содержании, которое когда-нибудь может пригодиться. Заучивание наизусть тут не поможет, ибо требуется разбирательство с содержанием, а не с формой.

- Если в ходе разбирательства с текстом что-то запомнено на внешнем носителе (экзокортекс/картотека/база данных: вынос мышления, в том числе используемой памяти и поиска в этой памяти ввне кортекса, то есть живого мозга), то это может быть найдено поиском и затем прочитано. Память и поиск перестают зависеть от прихотей природы, от аппаратных возможностей человеческого мозга. Блокнот для записей, персональная картотека, бортовой, или лабораторный, или производственный журнал (лог, записи происходящего – для памяти и последующего разбирательства) решают проблему.

- Память человечества – если какие-то мысли записаны в экзокортексе, то можно искать содержание во всех экзокортексах сразу. Есть разные варианты устройства такой системы: большие языковые модели (в том числе мультимодальные), например текстовая модель LLaMA содержит 220GB преднастроенной нейронной сети, в которые сжата текстовая информация объемом примерно в 1.4 триллиона токенов (частей слов, цифр, знаков препинания). Но поисковые системы типа Google дают и другой вариант: централизованно хранится индекс, по которому в датацентре довольно быстро находится адрес документов, наиболее вероятно содержащих нужные ключевые слова – и дальше можно уже обращаться к исходным документам. И это только два варианта устройства такой «памяти уровня человечества в целом». В бумажные времена такое тоже было, например, Библиотека Конгресса, только «память» эта не давала возможностей поиска нужной информации. Компьютеры решили и проблему хранения/памяти, и проблему поиска – причём поиск становится всё более изощрённым (например, поиск среди всех документов на всех языках сразу, а потом автоматический перевод найденного на язык, удобный человеку, которому эта информация потребовалась).

¹³⁸ https://en.wikipedia.org/wiki/Predictive_coding

Но мало владеть собственным умом и уметь его усиливать, надо ещё владеть телом: вокал/голос (выразительная речь) – это мышечная работа, работа на клавиатуре – это мышечная работа, сидеть восемь часов так, чтобы тело не отвлекало внимания на свои проблемы – это тоже мышечная работа, уметь расслабиться, чтобы считать невнятные кинестетические сигналы, которые даёт тело – это тоже мышечная работа. Уметь завязать себе шнурки в 60—80 лет – это мышечная работа, равно как и необходима мышечная работа, чтобы сохранить своё тело и мозг работоспособным на много лет. Это фитнес/fitness, как тренировка готовности к действию. А мышечная работа координируется мозгом, то есть тоже требует собранности и какой-то тренированности в управлении телом/body control.

Конечно, не нужно быть готовым в 21 веке к работе землекопом, к тяжёлому физическому труду «синих воротничков» (кнопки на станках нажимать сегодня не требует физической силы и выносливости), но поддержание не только собранности ума, но и собранности тела, готовности ума и тела к действию, а не к рассеянному «броуновскому» движению – это важная задача. Собранности ума и тела надо научить, причём это научение должно быть не только на личном уровне, но и уровне команды, коллектива как команды команд, организации как множества коллективов. Политтехнологи ставят задачи собранности для крупных сообществ и даже для обществ (все эти «всё для фронта, всё для победы» и «пятилетку в четыре года» – попытки удержать внимание многих людей на каких-то аспектах их жизни). Так что в обучении собранности нужно ещё разбираться: что надо удерживать во внимании, а на чём внимание надо наоборот, не удерживать, и предлагаемые объекты внимания «выкинуть из головы», хотя какие-то люди получают свою зарплату за попытки привлечь внимание личности, команды, коллектива, организации, сообщества/клуба, общества в целом. Надо быть разборчивым по поводу тех проектов, для которых нужно быть собранным.

Собранность как часть актёрского/ролевого мастерства

Мы уже рассмотрели ту часть, где актёров/акторов/агентов учат распознавать, в какую они попали пьесу/проект и какие проектные роли играют окружающие их участники проекта – для этого мы ввели идею **трудового кругозора**. В мире искусства это эквивалентно «насмотренности», начальному обучению нейронной сетки, чтобы дальше она могла воспользоваться и другими уровнями памяти и внимания, могла задействовать разные механизмы мышления и действия.

Но мы ещё не касались того, как человеку или даже подразделению самому научиться хорошо исполнять проектную/организационную/деятельностную/трудоую роль. Лидерство, которое изучают в системном менеджменте, учит «режиссёров» помогать «актёрам» хорошо исполнять их актёрские роли. Учительское мастерство – ровно то же самое: умение учителей помочь студенту стать мастером, хорошо исполнять какую-то роль, быть мастером какой-то практики. В искусстве мастерство лидера, мастерство учителя – это «режиссёрское мастерство». Актёрское/ролевое мастерство учит тому, как актёру/агенту/исполнителю роли выбирать роль (соглашаться, или не соглашаться на роль), входить в роль, выходить из роли, улучшать свою игру по роли. Нам нужно мастерство, аналогичное актёрскому.

Увы, знание того, как проходит обучение настоящих актёров актёрскому мастерству помогает, но не сильно. Актёров учат, например, изображать рассуждения своего персонажа, рассуждения роли – но нам-то **нужно учить не изображать прикладные рассуждения или даже мышление, а именно мыслить, решая проектные проблемы, которые нужно решать актёру согласно своей роли. Актёров учат эмоционально реагировать – но нам-то нужно содержательно реагировать, эмоции тут бантики сбоку (хотя их и нельзя игнорировать, но главным образом для того, чтобы замеченные эмоции держать под контролем, а не отпускать на самотёк). Агентам/акторам в проектных ролях нужно содержательно реагировать на проектные ситуации рассуждениями, в том числе мышлением (познавательными рассуждениями по трансдисциплинам интеллект-стека) и действием, а не эмоционально достоверно изображать своё отношение!**

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.