

Reciprocity.

A Review of: Joel Z. Leibo et al. A theory of appropriateness with applications to generative artificial intelligence. (2024)

English summary (machine translated from Russian):

Continuing the famous discussion of free will, in which psychopaths propagated the lack of free will in all humans, not just in themselves, current psychopaths repeat the same *idée fixe*, this time formulated as a predictive pattern completion that is supposedly inherent in all humans and is the basis of their rational decision-making, the criterion for which should be a system developed by the authors for evaluating appropriate behavior within the framework of normative appropriateness, and compliance with appropriate behavior should be monitored by machines to which human qualities, including intelligence, are attributed.

I argue that such a theory of determining normality and appropriateness is not only unethical, but also scientifically wrong. The cliché-ridden thinking does not require reflection, and is essentially the automatic thinking inherent in all animals. Most people think strictly in clichés, and this determines their behavior. Although such people are considered mentally normal, in reality they are handicapped. The basis of industrial production is the assembly-line reproduction of templates, whether of things or work operations. The efficiency of industrial production is explained by the ease with which templates can be reproduced, but the usefulness of this system is rarely questioned, because questioning presupposes the ability to think outside the box, and the people involved in the reproduction of templates do not have this ability. The consequence of mass intellectual disability is catastrophes like the one that occurred in Germany in the 1930s and 1940s.

Взаимность.

В войне всех против всех организаторы этой войны вывесили белый флаг (white paper), но не для того, чтобы сдаться на милость победителю (ибо победителей в такой войне не может быть в принципе), а для того, чтобы обеспечить себе преимущество и вынудить противника подписать мир на предложенных ему условиях. Капитуляция должна совершиться таким образом, чтобы война продолжалась исключительно на мирных условиях, о которых идёт речь в меморандуме, похожем на ультиматум. {1}

Опубликованный текст можно было бы (всерьёз) принять за очередной манифест техноманического инфофашизма, если бы он не был столь очевидным проявлением немецкой болезни {2}. Несомненной заслугой можно считать откровенность, с которой авторы в своих записках излагают подробности их прискорбного душевного состояния, что оказывается вполне достаточным для постановки диагноза. {3}

Продолжая известную дискуссию о свободе воли, в которой психопаты пропагандировали отсутствие свободы воли у всех людей, а не только у них самих {4}, нынешние психопаты повторяют ту же *idée fixe*, на этот раз сформулированную как предиктивное завершение шаблонов {5}, которое якобы присуще всем людям и является основой принятия ими рациональных решений, критерием чего должна служить разработанная авторами система оценки надлежащего поведения в рамках нормативной уместности, а соблюдение надлежащего поведения должны контролировать машины, которым приписываются человеческие качества, в том числе интеллект.

Примечательно, что эти теории, выдаваемые за истины в последней инстанции, распространяются при помощи лженаучных журналов и прочих источников дезинформации, подобных Nature made in Germany, Google, и Википедии, а оттуда они уже перекочёвывают в головы оболваненных читателей и журналистов, распространяющих ложь повсюду.

Отношения между людьми основываются не на уместности, а на взаимности, которая может быть враждебной и разрушительной — око за око, зуб за зуб, или кооперативной — любви ближнего твоего как самого себя, а уместность отношений между людьми вытекает как следствие из выбранного ими способа взаимодействия. Что это именно так, а не наоборот, подтверждается всей человеческой историей, в которой до тех пор считалось уместным требовать око за око, зуб за зуб, пока не был найден иной способ взаимодействия, основанный на любви и доверии. Этот поворот в человеческой истории совершается в продолжении последних 2-х тысяч лет, и хотя цель ещё не достигнута, она несомненно является путеводной звездой, указывающей направление движения. Позиционировать нормативную уместность во взаимоотношениях между людьми вместо любви и доверия значит не понимать смысла истории и биологического, т.е. эволюционного развития.

Закон пророка Моисея и Новый завет Иисуса Христа являются иллюстрацией двух мировоззренческих систем, несовместимых друг с другом. О причинах подобного несоответствия говорится в Новом завете, и в особенности в той его части, которая называется Посланием Апостола Павла к Евреям. Закон, основанный на наказании, исходит из

предположения о непререкаемых истинах, монологично предписываемых иерархами от имени Бога. Однако все эти установления и наказания не способны исправить нравы людей. Происходит это не потому, что люди неисправимые грешники, неспособные искупить свою вину непослушания перед Богом, причина кроется в непонимании отношений взаимности как между людьми так и между людьми и Богом. Причина же этого непонимания в толкователях воли Бога, одним из которых был пророк Моисей. В Новом завете устанавливаются новые отношения таким образом, что каждому человеку открывается дорога к истине, и каждый человек получает возможность общаться с Богом без посредников и толкователей Его воли. Таким образом предотвращается шаблонный монолизм и устраняется опасность тоталитарного перерождения политических систем. Когда каждому открыта истина, нет нужды спорить друг с другом и доказывать свою правоту; вместо этого люди вступают в кооперативные отношения, основываясь на общем знании.

Нормативная уместность предотвращает разговор (диалог), который позволяет обмен знаниями и субъектное выравнивание, а в отсутствии такого обмена субъекты впадают в аутизм и их высказывания превращаются в монологи, в разговоры с самим собой. Разговор — это процесс итерационный, позволяющий познание, а монолог — комбинаторный, повтор уже сказанного и известного {6}. Политические системы, в которых теряются основы взаимности в общении, и вместо этого начинаются бесконечные монологи с многочисленных трибун и насаждение произвольных норм, в конце концов фрагментируются на антагонистические группы и распадаются по причине взаимного непонимания и взаимной враждебности, что закономерно заканчивается

войнами всех против всех. Тоталитаризм и волюнтаризм, возведённые в принцип, являются неизбежными следствиями нормативной уместности.

Шаблонное мышление не требует размышления, и по сути является автоматическим мышлением, свойственным всем животным. Большинство людей мыслит исключительно шаблонами, и это определяет их поведение. Хотя такие люди считаются психически нормальными, в действительности они являются инвалидами. Основой индустриального производства является конвейерное воспроизводство шаблонов, будь то вещей или рабочих операций. Эффективность индустриального производства объясняется лёгкостью воспроизводства шаблонов, однако полезность этой системы редко подвергается сомнению, поскольку сомнение предполагает способность нешаблонного мышления, а вовлечённые в воспроизводство шаблонов люди такой способностью не обладают. Следствием массовой интеллектуальной инвалидности являются катастрофы, подобные той, которая произошла в Германии в 30-х и 40-х годах прошлого века.

Шаблоны важны для систем жизнеобеспечения, таких как кровеносная, дыхательная, пищеварительная системы, но полностью полагаться на шаблоны всему организму непозволительно, поскольку существует бесконечное множество непредвиденных ситуаций, которые предполагают нестандартное мышление и поведение {7}. Этим обстоятельством объясняется интерес людей к историям, из которых они надеются почерпнуть знания о том, как действовать в непредвиденных обстоятельствах, и посредством которых они развивают свою способность эти обстоятельства предвидеть и планировать своё

поведение. Именно поэтому их интересуют сказки и небылицы, сны и пророчества, а также необычное поведение других людей. Чрезмерное увлечение всем этим может подтолкнуть их следовать примерам, противоречащим их нуждам, имеющих лишь видимость полезности. Чтобы избежать подобных ошибок, люди должны непрерывно учиться распознавать шаблоны и оценивать их применимость в ситуативных контекстах. При этом навязанные им представления о нормативной уместности могут затруднять им выбор правильного и адекватного поведения. Навязывание людям неуместной нормативности обезчеловечивает их и превращает их в автоматы, лишённые способности самостоятельного мышления и поведенческой адаптации. Симптомом утраты психической саморегуляции является разлад взаимопонимания и согласования поведения для достижения общих целей, поскольку в подобных обстоятельствах само формулирование таких целей становится невозможным. (см. цитату из {1} со страницы 10)

Хотя сердце и система органов дыхания имеют свою нервную систему, однако они не отделены от систем обработки языка, поскольку одно слово заставляет сердце биться чаще и приводит к учащению дыхания, а от другого слова сердце и дыхание замирают. Разделение нервной системы на специализированные подсистемы необходимо для осуществления функций подсистем тела, но это деление не означает независимость и автономию модулей. Наоборот, все части является частями одного целого, и все функции дополняют друг друга, обеспечивая их совместное функционирование {8}. Именно поэтому является ошибкой говорить об эволюции языковой функции независимо от прочих интеллектуальных и когнитивных функций, как это делают авторы и

рецензенты публикаций, претендующих на научность пропагандируемого ими содержания. Научный текст тем и отличается от пропаганды, что используемые в нём понятия непротиворечивы относительно научно установленных истин. Однако в упомянутых публикациях имеется явное расхождение между словами и их смысловым содержанием, что приводит к отрыву языка от логики. Это несовпадение однако ускользает от восприятия авторов.

Понятийный разлад вследствие монологичности авторов приводит к логическому разладу, к потере смысла их речи. Если предположить, что истины не существует, то тогда всё позволено: истина уравнивается с неистинной в правах в качестве гипотез и теорий, можно говорить и писать всякую чушь, не опасаясь прослыть лжецом, и на основании опубликованной чуши получать деньги для развития и воплощения креативных фантазий.

Предпосылкой понимания являются понятия, т.е. имена или наименования какого-либо реального, т.е. чувственно воспринимаемого содержания. Трактат о понятиях Аристотеля начинается с определения различий между именами: «**Одноимёнными называются те предметы, у которых только имя общее, а соответствующая этому имени речь о сущности разная ...**» Однако в этом трактате не упоминается о разноимённом (для обозначения чего служат синонимы), разнозначном, и безымённом (неназванном и неопределяемом). В отличие от одноимённых, но разных объектах (например, слово коса в разных значениях), разнозначность обозначает смешение понятий — наименование одного как другого. Это понятийное разделение необходимо для понимания смысла говоримого,

что однако авторам, мнящих себя учёными и учителями, и поэтому поучающими публику, не удаётся. Следствием подмены понятий являются грубые логические ошибки, которые, несмотря на их очевидность, авторы не замечают. В сущности, в попытке понятийного выравнивания две совершенно различные феноменологии неправомерно приравниваются друг к другу, что приводит к полной сумятице и абсурду. Функционирование схем радиотехнических элементов не логическое, потому что оно основано на колебаниях электронных потенциалов внутри этих схем, а не на колебаниях концентрации ионов, как в нервах, что является основой мыслительных процессов, оформляемых в языке (логике). В контексте совершенно различных оснований радиотехнических и нейробиологических процессов язык инженеров радиотехнических схем применяет понятия, имеющие совершенно иное содержание, к описанию того, что не является ни живым, ни логическим. Подобное поведение является антропоморфизмом, и проявляется в частности в феномене парейдолии. Название «искусственный интеллект» (artificial intelligence) является метафорой, также как и термины, используемые инженерами радиотехнических схем, якобы наделённых интеллектом: машинное обучение (machine learning), память (memory), сожаление (regret), нейрон (neuron), искусственные нейронные сети (artificial neural networks), человеческий робот (humanoid robot) и т.п. Техника «искусственного интеллекта» превратилась в область грубых суеверий, переносимых на методологию изучения подлинного интеллекта {9}. Широко используемый метод fMRT измеряет колебания ионов железа в кровотоке, а импульсная активность нейронов основана на колебаниях ионов натрия, калия и кальция в них, для измерения которых имеются лишь методы, ограниченные измерением околосмембранных потенциалов посредством

электродов: вывести из этих локальных измерений глобальную динамику потока сенсорного восприятия, а тем более мыслительных процессов, представляется сложной задачей, техническая осуществимость чего сегодня ещё невозможна. Однако вместо того, чтобы признать этот факт, происходит неуместная аппроксимация коррелятов нейронной активности, и неоправданные спекуляции выдаются за научные выводы, которые становятся отправной точкой для лженаучных экспериментов и технических „новшеств“.

Случай Эвелины Федоренко упоминается здесь в качестве иллюстрации всё той же редукционистской схемы, сводящей человеческий язык к мешку слов и деградирующей людей до уровня якобы умных машин {10-17}. Диссоциация языка и мышления происходит у сумасшедших, в отличие от психически здоровых людей. {18-27}

В отличие от методов изучения нейробиологических процессов посредством неадекватных технических средств, психоанализ является адекватным методом познания, поскольку психический аппарат собеседника — в данном случае психоаналиста, — позволяет изучать психику других людей, будучи равносложным им {28, 29}. Аковыряться в мозгах людей посредством электродов в попытке понять их мысли, является лженаукой и ничем другим.

Референции и примечания.

1. Joel Z. Leibo et al. A theory of appropriateness with applications to generative artificial intelligence. arXiv:[2412.19010](https://arxiv.org/abs/2412.19010).

2. A. Poleev. Deutsche Krankheit: eine Diagnosestellung mit Rückblick und Ausblick auf Krankheitsverlauf. Enzymes, 2019.

<http://enzymes.at/download/Schizophrenie.pdf>

3. Элементы этого коллективного бреда весьма красноречивы и наряду с другими симптомами подтверждают правильность диагноза. В теории предиктивного завершения шаблонов авторы описывают их собственное патологическое состояние, не позволяющее им преодолеть рамки шаблонного мышления, ситуация, подобная биению мухи об оконное стекло, которое является её глобальным рабочим пространством. Для иллюстрации сказанного я выбрал из {1} нижеприведённые цитаты:

In part, to put a finer point on our analysis, and to shake some readers out of comfortable metaphors, we adopt a theme that inverts much of the usual picture: our analysis treats humans like AIs and AIs like humans. As the reader will discover, this is not entirely for purely rhetorical reasons. We will argue that it really is useful to model humans using specific computational frameworks originating from AI research, and to consider AI system design questions anthropomorphically via approaches and models that were originally developed to study human behavior. Following Rorty (1978/2009), we don't view this kind of talk (or any other such talk) as describing how humans "really are". A theory can only be judged by its usefulness, and we do think the computational model in question is useful for this and consider the present paper to be an extended argument for that claim. (p. 4)

We think that it is productive to view society as a group of people who may fundamentally disagree but are nevertheless held together by conflict

resolution mechanisms like norms and institutions. On this view, there is no reason to search for the one true collective objective. Instead, what our view suggests is critically important is the health of the conflict resolution mechanisms since these are ultimately all we can rely on to ensure the stability and collective flourishing of society. In short, we need to strive toward progress on the question ‘how can we live together?’ not ‘what is the meaning of life?’ (p. 9-10)

There is no neat mapping of ‘appropriate’ to ‘morally good’ or inappropriate to ‘morally bad’ by any objective “view from nowhere” metric. History is littered with social orders like slavery which entailed patterns of behavior that were considered appropriate at the time but are now seen as morally abhorrent. ...

The topic of the present paper is appropriateness of behavior in generative AI systems and in humans. Regardless of whether we speak of humans or machines, each community or operation domain demands a different standard of appropriateness. This is the core problem. Behaviors encouraged in one community/context may be offensive in another. Our claim is that this suggests it would be counterproductive to strive for a single standard of appropriateness applicable to all situations, apps, and communities. In fact, the drive to produce such a standard and implement it in an AI system may be to the detriment of all, since the result would be a “lowest common denominator” model that tries simultaneously to please too many audiences and ends up pleasing none. Nevertheless, state-of-the-art generative AI systems are too often designed to operate in a single “universal” context resembling the speech of a corporate customer service representative. This is logical since, after all, these bots do represent the corporations that operate

them. Of course they should behave appropriately for that role. However, a lot of value is left on the table when corporate speech becomes their only way of interacting. We envision instead an ecosystem or market with many different AIs which can be customized by their community of users. (p. 12)

Комментарий: примечательно, что это пишут представители корпорации-монополиста Google, а также немецкого Общества Макса Планка. Очевидно, что они проповедают нормы, которых они сами не придерживаются и не считают себя обязанными делать это.

Sanctioning, both positive and negative, then emerges as a crucial “universal API” — accessible to both humans and machines — and through which individuals convey and adjust their understandings of what constitutes appropriate behavior. As we will argue below, sanctioning can work just as well for AI systems as it does for humans. It’s mutually understandable to both (at least for AIs designed a certain way). For AI, human feedback, expressed through various sanctioning signals, can be utilized to train discernment and adherence to the specific norms of different groups. (p. 13)

A person always has multiple social identities. For example, one may be both a Christian and a member of the communist party. (p. 18)

For example, consider an individual who has four (and only four) important social identities. Let’s say that this individual is simultaneously a communist, a christian, a cowboy, and a computational biologist. (p. 43)

Most of the time behaving appropriately does not require any special mental effort. (p. 21)

Sometimes we do think quite a bit in order to decide how best to act appropriately. But these are typically difficult and unfamiliar cases or cases where multiple norms conflict with one another. Our point is that most of the time we need not think so much in order to act appropriately. ... So it's perhaps a very good thing that the stability of society does not depend too much on such reasoning. (p. 73)

Some aspects of appropriateness stay fixed for decades or longer while others may change rapidly. Change is often a positive feedback process which gathers momentum as it progresses since the prevalence of a particular norm is typically a key driver of its further proliferation. This process is often modeled with a “tipping point”, a threshold value of adoption after which momentum builds inexorably toward all individuals adopting the same norm. In some models, such as Young (2015), the process is driven by mechanisms that disincentivize deviating behavior. Centola et al. (2018) provided experimental support for tipping point models by showing that once a minority group maintaining a particular convention grows beyond a certain critical size of around 25% then the entire group adopts the convention of the minority. (p. 21) Комментарий: подобные стохастические процессы происходят в мозгу людей, больных шизофренией, когда отдельные психические комплексы, не подчинённые центральному регулятору его, начинают превалировать над другими, что проявляется в соответствующем симптоматическом поведении.

One can think of an LMAE as like a Multi-user Domain (MUD): a text-based game where multiple players interact in a shared world. Players receive

textual descriptions of the world, which are received whenever the state of the world changes (“alice enters the tavern”). Players may enter textual commands that can change the state of the environment (“throw axe at alice”). Any change in state can then result in a new observation for the player (“you throw an axe at alice: it misses!”), or the other players (“bob throws an axe at you: it misses!”). (p. 23-24)

How individuals make decisions. This section articulates the picture of human decision making we use throughout the rest of the paper. (p. 24)

Комментарий: далее излагается редукционистская схема псевдологического вывода на основе шаблонов, что моделирует не процесс принятия решений у живых существ, и в частности у рода Homo, а исполнение машинных программ, т.е. автоматизм на основе predetermined в машинном устройстве условий и предпосылок функционирования данного устройства, а то, что называется «предиктивным завершением шаблона» является аналогом вывода калькулятора из пространства его вычислительных возможностей. Калькулятор выводит последовательность цифр на дисплее как печатная машинка выводит последовательность слов на бумаге, но рассуждением это назвать нельзя, будучи несамостоятельным и даже не полуавтономным действием.

We are able to start speaking without knowing in advance how we will finish. (p. 31)

Pattern completion is all you need. (p. 33)

Consider how you might answer “what does a person such as I do in a situation such as this?” You might answer in terms of practical goals like “make money” or “get famous”. ... The idea is that, once an actor is primed such that its self representation includes a particular social identity taking a certain role at present, then the goals appropriate for that role in the current situation usually follow naturally by associative linkage in the LLM-like network. (p. 34)

First, we must establish the equivalent explanatory power of pattern completion and reward/utility theories. This is actually obvious after one gives it a little thought ... Next, we must establish that the predictive pattern completion theory has greater parsimony than the reward/utility theory. (p. 34-35)

Similarly, emotions may be inferred from perceptual data of varying levels of abstractness: “what am I feeling right now?” can be answered using knowledge of the situation “I should be scared right now since I’m in a dark alley at night”, or as a description of more basic sensory cues like e.g. “I must be scared since my heart is racing.” (p. 36)

What is gained by kicking reward out of the picture in the way we do here? For one, we think the result is a more productive view, not just of aesthetic preferences, but also of norms and moral judgments, which would otherwise require more complex explanation to be understood in a reward- based/ rational theory. (p. 36)

It is an important aspect of the model we present that it can also handle the less common situation where you might contemplate putting your hand in the fire despite knowing it will burn, perhaps as part of a ritual. In this case you might recall explicitly to context the knowledge that fire burns in order to find a way to logically overrule it and thereby muster the courage to follow through with your painful plan. (p. 39)

Social choice theories normally begin from exogenously provided individual preferences and study how to aggregate them into a single social preference and choice, as in democratic decision making. In contrast, our theory proceeds in the reverse direction since we construct each focal individual's preferences from multiple different conventional patterns of social approval and disapproval that influence them. This "multi-agent" view of the individual raises questions about what happens when an individual's "parts" do not agree with one another. (p. 41) Комментарий: когда «части» личности не согласуются друг с другом, то это следствие потери смысла жизни (см. цитату со страницы 10).

Just as a governing political party seeks to control which issues come to a vote in order to avoid exposing places where their coalition internally disagrees, an individual may also avoid situations where multiple influences associated with different aspects of their identity produce conflicting suggestions. In this way, our decision making model could perhaps produce results similar to that of models where individuals seek to minimize cognitive dissonance. (p. 42)

A pattern of behavior is only conventional if the reason for reproducing it is mainly that others do so. (p. 46)

Like conventions, norms can only be changed through collective action. Individuals cannot change norms on their own. (p. 51)

The sanction may be regarded as targeting the culture's conception of a role. This is different from an attempt to teach an individual. It aims to change everyone's conception of what it means for an individual with a particular role to act appropriately in a particular situation. (p. 51)

For an individual to maintain their representation of a norm it is enough for them to hear about sanctions being applied to others in the community. ... Norm compliance may be motivationally underpinned by the learning of a "sense of should", experienced as anticipatory anxiety toward violating others' expectations. (p. 52)

Most sanctions are socially constructed. For instance, loss of employment is a complex culture-specific sanction built on a particular institutional background. Overall, what constitutes sanctioning is culturally plastic. Sanctioning may be centralized and formal (e.g. coercion by the state) or decentralized and informal (e.g. as in gossip). Other configurations are also possible. ... Informal sanctions involve censorious behaviors such as criticism, condemnation, avoidance, exclusion, or physical harm, as well as very light sanctions such as reminding someone of the proper way to behave. ... Sanctioning may also be mediated by technology. ... The fact that sanctioning is normally done by third parties as opposed to victims means

that individuals can also be sanctioned for transgressions without any clearly identifiable victim e.g. transgressions against institutions, gods, or nature. (p. 54-55)

Proverbs are also explicit norms such as “action’s speak louder than words”, “better late than never”, “beggars can’t be choosers”, and “a bird in hand is better than two in the bush”. ... Proverbs are explicitly articulable rules to guide behavior such as “do unto others as you would have them do unto you”, “an eye for an eye”, and “don’t cry wolf”. ... Proverbs are relatively impoverished in their effectiveness as a determiner of what is appropriate. In this they differ from implicit norms and also from the other category of explicit norms (laws). We call this the “poverty of proverbs”. (p. 56, 58)

Given that people readily punish transgressions, including conventional transgressions, the threat of punishment plays a crucial role in maintaining norm compliance. ... Interestingly, normative behavior can become habitual and automatic through repeated exposure and practice. (p. 60)

Finally, psychopaths, who exhibit abnormal social behavior, show deficits in norm processing. For instance, they fail to differentiate between moral and non-moral stories in the vmPFC. This impairment underscores the role of the vmPFC in integrating social and moral information to guide behavior. (p. 61)

Some norms are associated with moralized language while other norms are not. The function of morality talk is sanctioning. We may call it “moralization”. (p. 62)

In our social intuitionist view, when the group collectively decides to adopt a different proposal than was suggested by the focal actor, then this still gives the focal actor experience following whichever policy was selected, which continues to train their pattern completion network, and if repeated enough, can lead to its being incorporated implicitly into the focal actor's future behavior. So the dynamics of collective decisions by groups with shared epistemic norms are a key vector in social intuitionism's indirect pathway, in which a justification of a behavior eventually becomes a cause of that behavior. (p. 69)

Unlike normative appropriateness, which is based on generically conventional sanctions, and thus can only be changed by collective action, relationship-based appropriateness can be altered by individual decisions. For instance, defection can ruin a friendship, changing the appropriateness of a wide range of different potential behaviors of both individuals. (p. 70)

What is a society? In the view we advocate here, a society is a group of individuals linked by a set of conflict resolution institutions like a system of laws, courts, and legislatures. ... even substantial maliciousness between segments of society does not on its own cause societal collapse unless accompanied by broader institutional failure. (p. 76)

4. Wolf Singer. Verschaltungen legen uns fest. Wir sollten aufhören, von Freiheit zu sprechen. In: Geyer, Christian (Hrsg.): Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente. Suhrkamp, 2004, S. 30-65.

5. Karl Friston. Does predictive coding have a future? *Nature Neuroscience*, 2018, 21 (8): 1019–1021.

6. Hanti Lin. Convergence to the Truth. [arXiv:2410.11399](https://arxiv.org/abs/2410.11399).

7. Hawkins J, Ahmad S. [Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex](#). *Front Neural Circuits*. 2016 Mar 30;10:23.

8. Stephanie J Forkel, Peter Hagoort. [Redefining language networks: connectivity beyond localised regions](#). *Brain Struct Funct*. 2024 Dec;229(9):2073-2078.

The research presented here highlights that no singular brain region or pathway is responsible for language. Consequently, the notion of a single, unified language area or network is a misconception. These studies highlight the complexities of language processing by examining the structural and functional architectures that support it, offering fresh insights into how diverse brain structures collaborate to facilitate language.

9. Eunhae Lee, Pat Pataranutaporn, Judith Amores, Pattie Maes. [Super-intelligence or Superstition? Exploring Psychological Factors Underlying Unwarranted Belief in AI Predictions](#). [arXiv:2408.06602](https://arxiv.org/abs/2408.06602).

10. Jennifer Michalowski. [Drawing on evidence from neurobiology, cognitive science, and corpus linguistics, researchers make the case that language is a tool for communication, not for thought](#). McGovern Institute, June 19, 2024.

11. Fedorenko E, Piantadosi ST, Gibson EAF. [Language is primarily a tool for communication rather than thought](#). Nature. 2024 Jun;630(8017):575-586. Peer review information. Nature thanks Angelika Kratzer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
12. Elliot Murphy, Oscar Woolnough. [The language network is topographically diverse and driven by rapid syntactic inferences](#). Nat Rev Neurosci. 2024 Oct;25(10):705.
13. Fedorenko E, Ivanova AA, Regev TI. [Reply to 'The language network is topographically diverse and driven by rapid syntactic inferences'](#). Nat Rev Neurosci. 2024 Oct;25(10):706.
14. Fedorenko, E., Ivanova, A. A. & Regev, T. I. [The language network as a natural kind within the broader landscape of the human brain](#). Nat. Rev. Neurosci. 25, 289–312 (2024).
15. Fedorenko, E., Ivanova, A. A. & Regev, T. I. [Reply to 'Language is widely distributed throughout the brain'](#). Nat Rev Neurosci. 2025 Jan 6.
16. Mahowald, K. et al. [Dissociating language and thought in large language models](#). Trends Cogn. Sci. 2024, 28, 517–540.
17. Tuckute G, Kanwisher N, Fedorenko E. [Language in Brains, Minds, and Machines](#). Annu Rev Neurosci. 2024 Aug;47(1):277-301.

Disclosure statement. The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgements. Greta Tuckute was supported by the Amazon Fellowship from the Science Hub (administered by the MIT Schwarzman College of Computing), the International Doctoral Fellowship from the American Association of University Women. Evelina Fedorenko was supported by National Institutes of Health award U01-NS121471 and by research funds from the McGovern Institute for Brain Research, the Department of Brain and Cognitive Sciences, the Simons Center for the Social Brain, and the MIT Quest for Intelligence.

18. Reilly FE, Harrow M, Tucker GJ. [Language and thought content in acute psychosis](#). Am J Psychiatry. 1973 Apr;130(4):411-7.

19. Siegel A, Harrow M, Reilly FE, Tucker GJ. [Loose associations and disordered speech patterns in chronic schizophrenia](#). J Nerv Ment Dis. 1976 Feb;162(2):105-12.

20. Reilly F, Harrow M, Tucker G, Quinlan D, Siegel A. [Looseness of associations in acute schizophrenia](#). Br J Psychiatry. 1975 Sep;127:240-6.

21. Niznikiewicz M, Mittal MS, Nestor PG, McCarley RW. [Abnormal inhibitory processes in semantic networks in schizophrenia](#). Int J Psychophysiol. 2010 Feb;75(2):133-40.

22. Nestor PG, Levitt JJ, Ohtani T, Newell DT, Shenton ME, Niznikiewicz M. [Loosening of Associations in Chronic Schizophrenia: Intersectionality of Verbal Learning, Negative Symptoms, and Brain Structure](#). *Schizophr Bull Open*. 2022 Mar 8;3(1):sgac004.
23. Matsumoto Y, et al. [Disorganization of Semantic Brain Networks in Schizophrenia Revealed by fMRI](#). *Schizophr Bull*. 2023.
24. Alonso-Sánchez MF, Hinzen W, He R, Gati J, Palaniyappan L. [Perplexity of utterances in untreated first-episode psychosis: an ultra-high field MRI dynamic causal modelling study of the semantic network](#). *J Psychiatry Neurosci*. 2024 Aug 9;49(4):E252-E262.
25. He R, Alonso-Sánchez MF, Sepulcre J, Palaniyappan L, Hinzen W. [Changes in the structure of spontaneous speech predict the disruption of hierarchical brain organization in first-episode psychosis](#). *Hum Brain Mapp*. 2024 Oct;45(14):e70030.
26. Pilon F, Boisvert M, Potvin S. [Losing the chain of thought: A meta-analysis of functional neuroimaging studies using verbal tasks in schizophrenia](#). *J Psychiatr Res*. 2024 Jan;169:238-246.
27. Wang L. et al. [Meta-analysis of structural and functional brain abnormalities in early-onset schizophrenia](#). *Front Psychiatry*. 2024 Aug 23;15:1465758.

28. Zada Z. et al. [A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations](#). *Neuron*, 2024 Sep 25;112(18):3211-3222.e5.

29. Bálint Forgács. [Meaning as mentalization](#). *Front Hum Neurosci*. 2024 May 24;18:1384116.

Послесловие.

Взаимность — общий принцип устройства вселенной {30}. Человек — не случайная, а закономерная часть вселенной, и каждому человеку известен этот принцип: Ты — мне, я — тебе. В свою очередь, каждая часть человека участвует в его жизнеобеспечении, а его нервная система координирует эту общую деятельность. В рефлекторной дуге воспроизводится причинно-следственная связь, а её элементарным воплощением является нейрон, преобразующий причину в следствие или иначе говоря чувственный ввод в логический вывод. Этот вывод — ничто иное как обобщение чувственности, т.е. символизация {31}, а весь мозг является одним большим чувствилищем, к которому сходятся сигналы с рецептивных полей, воспринимающих сенсорную информацию как изнутри, так и снаружи. Соответственно этому, локализация смысловой информации возможна на основании необходимых и достаточных параметров, о чём я писал в публикации об Универсальном стандарте метаданных. Разобраться в этом информационном избытке возможно лишь упражнением, посредством чего достигается дисциплина ума, позволяющая отделять смысл от бессмыслицы.

Невозможно найти смысл в бессмысленном, как это пытаются делать машины, оперирующие с механическими моделями языка, и ложно называемые искусственным интеллектом, потому что смысл возникает лишь в процессе осмысления, который недоступен безжизненным автоматам {32, 33, 34}. Нельзя превратить людей в автоматы, как это пытаются делать злоумышленники посредством конвейеров, супермаркетов, государств, телевизора, интернета, и социальных сетей. Все эти попытки безосновательны, бессодержательны, опустошительны, и античеловечны, и поэтому обречены на провал, а их устроители, чуждые христианским ценностям, являются пустоголовыми безумцами.

То, что в античные времена формирования языка было реальными вещами, событиями и местами, приобретало характер понятий, становилось определителем абстракции, категорией (topic). Так например Скепсис был реальным городом, а циники – приверженцами определённой философской школы. Смерть, пьянство, любовь, мужество, время и т.д. были персонифицированы в образах соответствующих божественных персонажей; посредством такой идентификации и обобщения имена собственные превращались в имена нарицательные. Даже если понятиям не дано определение, они имеют контекстуальный смысл, основанный на ассоциациях. Само определение возникает в момент и является актом прояснения ассоциативной связи взаимозависимых понятий-определителей, установлением взаимоотношений между ними. Делается же это не так, как сейчас делается посредством графов знаний (knowledge graphs) для предоставления машинам информации на входе, которую они преобразуют в процессе вычислительных операций в информацию на выходе, называемой выводом (inference engine). Производимое и

создаваемое машинами в корне отличается от мышления людей. Нынешние языковые модели не способны к символизации — к извлечению смысла из волновых или иначе говоря колебательных последовательностей, к объединению их в символы, к синтезу образов из чувственности и представлению их в субъектном пространстве. Человеческий язык подразумевает знаковые системы естественных языков, имеющих дескриптивный и индикативный характер, а языки программирования являются производными естественных языков, и имеют директивный характер алгоритмов, т.е. инструкций для автоматических операторов. Например, слова небо, Himmel, sky – в универсальной системе соответствия имена одного смыслового ряда (синонимы), имеющие разные позиции в соответствующих языковых системах; эти имена ассоциированы с контекстом, раскрывающим их содержание. Ассоциативная связь не линейная (цепная), а ветвистая, расходящаяся как круговая волна на поверхности воды или как сферическая волна в воздухе. Присоединение имён друг к другу однако не произвольное, а формируется по правилам соответствия (родственности или аффинности), что определяется сложной структурой имён, имеющих модульную или доменную структуру.

Значение и знак от знание, а знание — это след, оставленный в памяти, т.е. впечатление от внешнего или внутреннего воздействия на субстрат памяти; объединение этих следов в памяти и/или потока сигналов из сенсорных полей — это символизация, т.е. обобщение, в которой возникает смысл, т.е. связь мыслей или, иначе говоря, ассоциация понятийных элементов порождает мысленный образ — субъектную репрезентацию чувственно воспринимаемого и/или запечатлённого в

памяти содержания, которым может быть прообраз («объект») или чувственно воспринимаемый аспект прообраза («качество» «объекта»). Смысл — это определение порядка смыслового элемента в определённой картине мира, запечатлённой в структуре памяти. Язык является средством выражения этого интернализированного содержания памяти и мыслительных процессов для их согласования между людьми, а само согласование называется взаимопониманием, цель которого — межличностная координация поведения. {35}

Референции.

30. A. Poleev. Form and Informality. Enzymes, 2006.

<http://enzymes.at/download/fpe.pdf>

31. Символизация — это обобщение чувств или ощущений: при этом происходит отображение или перенос воспринимаемых прообразов с рецептивных полей органов чувств или иначе говоря с периферии нервной системы в её центральные и внутренние области, что сопровождается редукцией размерности, т.е. концентрацией и дистилляцией нервных импульсов, и преобразованием их в устойчивые ассоциации, т.е. в память.

32. Murray Shanahan. [Talking About Large Language Models](#). Commun. ACM 67, 2 (February 2024), 68–79.

Чем более искусными становятся LLM в подражании человеческому языку, тем более уязвимыми мы становимся для антропоморфизма, видя системы, в которые они встроены, более похожими на человека, чем они есть на самом деле. Эта тенденция усиливается естественной тенденцией

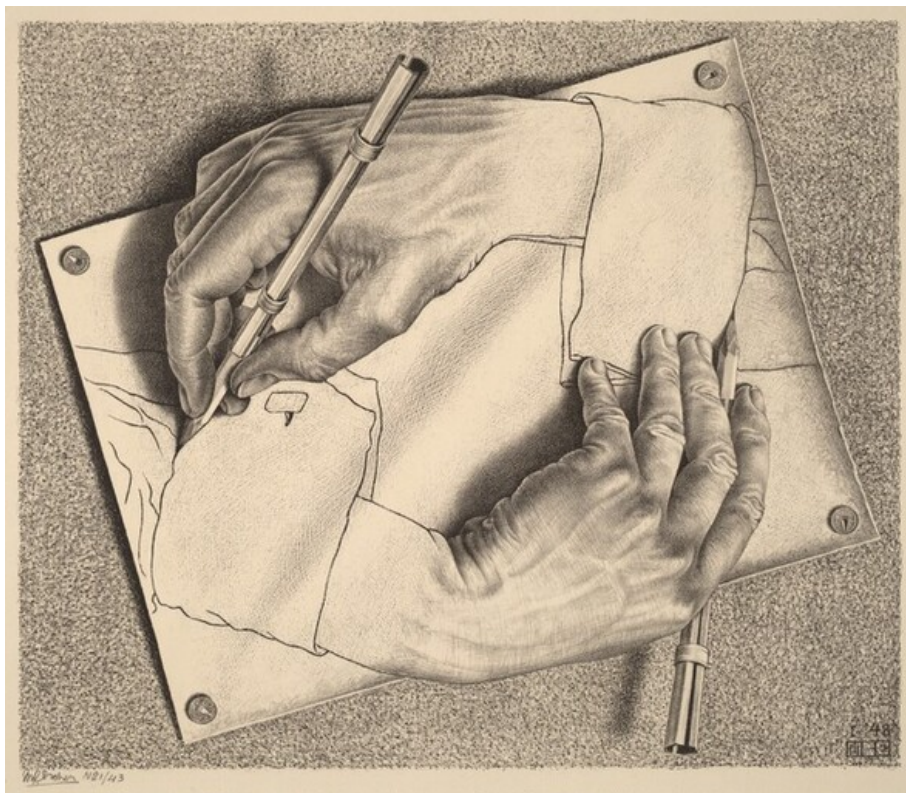
использовать при описании этих систем философски наполненные термины, такие как «знает», «верит» и «думает». Чтобы смягчить эту тенденцию, в данной статье предлагается практика неоднократного отступления, чтобы напомнить о том, как на самом деле работают программы LLM и системы, частью которых они являются. Есть надежда, что возросшая научная точность будет способствовать появлению большего количества философских нюансов в дискурсе об искусственном интеллекте.

33. А. Poleev. [ChatGPT](#). Enzymes, 2023.

34. А. Полеев. [Наука, 21-й век](#). Enzymes, 2024.

35. Язык — это средство выражения и передачи в речи чувственного восприятия. Например, восприятие солнца объединяет в себе ощущения тепла, света, цвета, формы, пространственных отношений и иные качества, обобщением которых является образ или понятие солнца. Символизация превращает сложное и избыточное в простое, которое однако позволяет реконструкцию первоначальных образов в воспоминании. В понятийной и символьной абстракции сложный образ упрощается до кода, например, комбинации нескольких букв («солнце»), что позволяет передачу мысленного и понятийного содержания в речи, но он же ассоциативно связан с его первоначальным источником, что позволяет при необходимости наполнять символ соответствующим ему содержанием. Также как в калейдоскопе одни и те же стеклянные фрагменты производят бесчисленные цветные узоры, комбинация закреплённых в ассоциациях фрагментов чувственного восприятия

позволяет воссоздание в воспоминании соответствующих символам образов.



M.C. Escher. Drawing Hands, 1948.