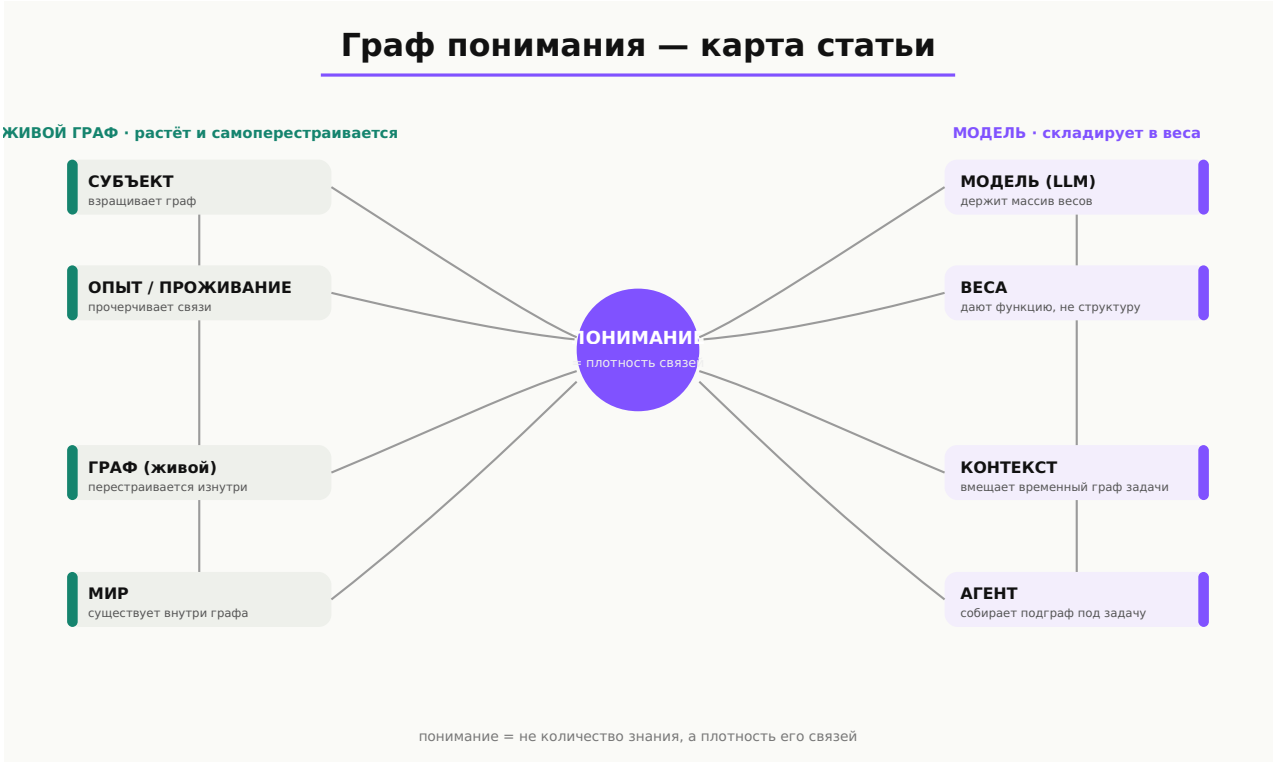
Граф против весов

Два способа поместить мир в голову — и почему агент меняет правила игры

Лонгрид · человеческим языком

> У меня есть подозрение, что человеческий мозг и современная нейросеть отвечают на один и тот же вопрос — «как поместить мир в голову» — двумя противоположными способами. Один врастает в знание, как дерево врастает в почву. Другой складирует знание, как амбар складирует зерно. Результат на первый взгляд похож: и там, и там «много знают». Но разница не в том, сколько знаешь, а в том, чем ты этому знанию являешься. Давайте разберёмся, не сбегая в терминологию.



Вся статья на одном листе — карта связей, по которой пойдём

Телега и лошадь

Есть старая картинка про неправильную упряжь: когда телегу ставят впереди лошади. Ехать так нельзя — телега не тянет, её можно только толкать вечно, упёршись в зад подводы. Нормально всё устроено иначе: лошадь идёт впереди и тянет телегу за собой — двигатель ведёт груз, причина двигает следствие.

Я всё чаще ловлю себя на том, что эта картинка всплывает сама собой, когда говорят об искусственном интеллекте. Потому что способ, которым мы сейчас учим большие модели, по устройству своему — это телега, поставленная впереди лошади. И чтобы это почувствовать, не нужно быть программистом. Достаточно спросить себя: а что вообще значит «понимать»?

Как человек растёт во знание

Представьте ребёнка. Первые годы он не «загружает» мир в себя, как файл в папку. Он его проживает. Трогает, роняет, обжигается, спрашивает «почему», снова трогает. Каждое повторение не записывается отдельной строкой в базу — оно меняет ткань того, чем он это представление держит. Сегодняшняя ошибка присоединяется к вчерашней, обе — к «что такое горячо», а «горячо» связано с «крик», «обжёгся», «больно», «не трогай», «я же говорил».

Нарисуйте это на бумаге — получится не список, а **граф**: узлы-понятия и связи-отношения. Чем дольше ребёнок живёт и учится, тем больше узлов и — что важнее — тем плотнее они связаны между собой. Не «он знает двести слов», а «в его голове слово связано с ощущением, картинкой, историей, запахом, чужим лицом, собственной болью». Каждое понятие висит на сотнях нитей, и каждая нить — это прожитый опыт.

И тут начинается самое важное. Потому что именно **плотность этих связей и есть понимание.** Не количество узлов, не объём словаря, а степень связности. «Понимаю» означает не «могу повторить», а «могу увидеть, как это стыкуется с десятком других вещей, и действовать». Миропонимание — это не факт, это структура.

> Плотность связей определяет миропонимание. Плотность создаёт свою реальность.

Вот ключ: для человека мир существует **только внутри его графа**. Мы не имеем доступа к «миру как он есть» — мы имеем доступ к топографии собственных связей. Двое одинаково одарённых людей, выросших в разных средах, живут в слегка разных мирах, потому что их графы связаны по-разному. Отсюда вечные споры «как можно так не понимать очевидного»: то, что для одного — дорога, для другого — просто чужой рисунок на карте, по которой он никогда не ходил.

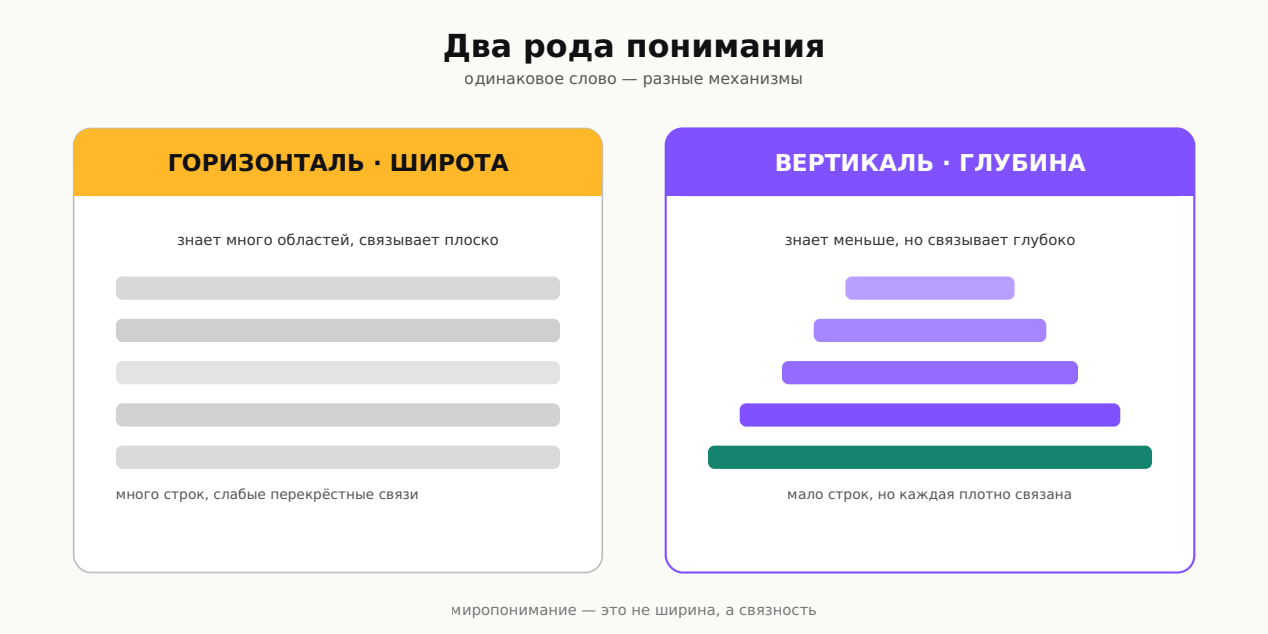
Два рода понимания

Стоит замедлиться и развести две вещи, которые мы называем одним словом.

Горизонтальное понимание — это широта. Человек или система знает много областей: техника, медицина, история, юриспруденция. Но знает плоско: по

каждой теме способен назвать факты, а глубоко связать и применить к живой ситуации — не всегда. Это информированность.

Вертикальное понимание — это глубина. Узкая область, но укоренённая: каждое понятие плотно связано с опытом, ошибками, следствиями, чужими реакциями. Это то, что мы называем мастерством и чутьём.



Оба нужны. Но важно понять, что это разные механизмы, а не степени одного. И именно здесь возникает путаница, когда мы сравниваем человека и нейросеть.

Как нейросеть складирует знание

Большая языковая модель — это не дерево и не граф изначально. Это **огромный массив чисел — веса**, которые устроены совсем иначе: в них нет «понятий» и «связей» как отдельных объектов. Есть лишь коэффициенты, настроенные так, чтобы при любом наборе входных сигналов выдавать правдоподобный выход. Учится это дело путём **корректировки весов**: показали пример — она ошиблась — поправили числа — показали ещё миллион — числа поехали в сторону, где ошибок меньше.

Здесь и спрятана телега, поставленная впереди лошади. Человек сначала строит связную картину и из неё действует. Модель сначала получает функцию и из функции выдаёт ответ, а связность ей приходится выуживать как побочный эффект. Мы скормили ей телегу опыта и надеемся, что она сама соберётся в лошадь.

Парадокс в том, что модели известно невероятно много — куда больше любого человека. Но знание это горизонтальное: широкое и плоское. А человеческое — вертикальное: пусть уже, но глубоко укоренённое и плотно связанное в работающий механизм.

Почему граф тянет телегу, а веса — позади

Разберу это медленно, потому что здесь стержень статьи.

Когда человек сталкивается с новой задачей, он не пересчитывает всё заново. Он спускается в свой граф и ищет самый плотный маршрут: вот эта ситуация выглядит как та, что была в прошлый раз; вот эти признаки цепляются за «осторожно»; а «осторожно» тянет за собой «проверь края». Понимание — это способность ходить по своим связям, и с каждым шагом граф уплотняется: сегодняшний случай добавил ниточку между «похоже на прошлый раз» и «всегда проверяй края».

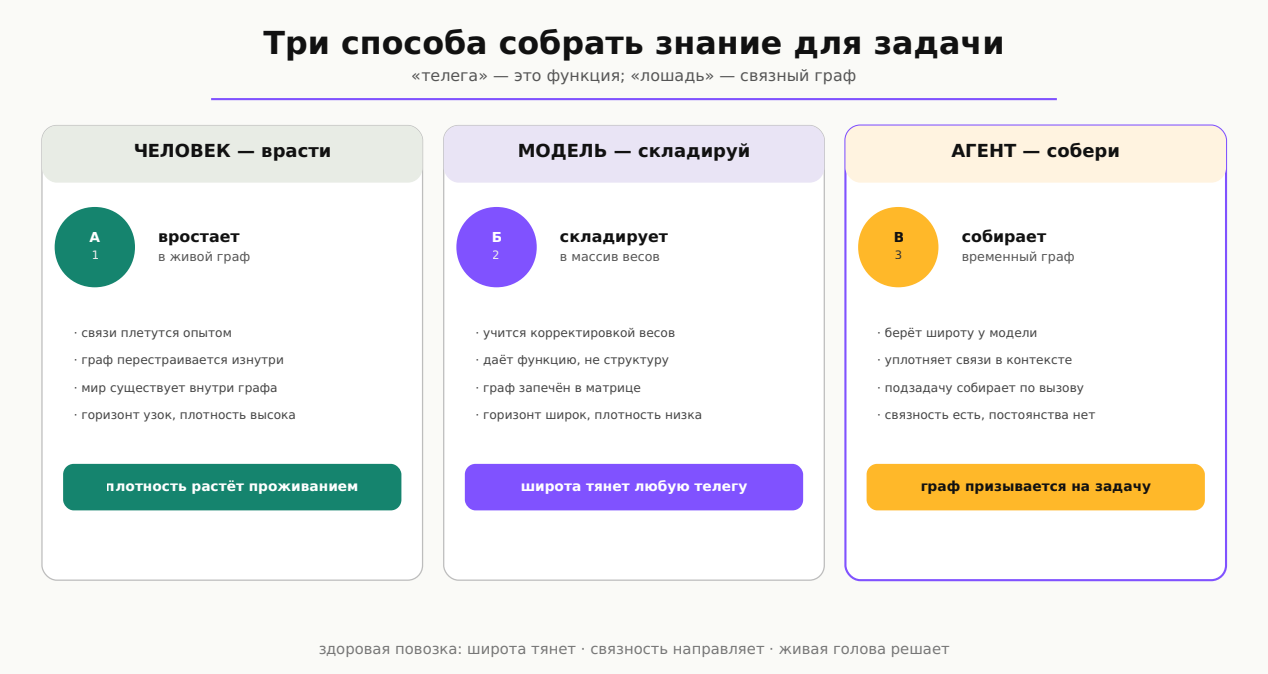
Теперь представьте, что этой способности — ходить по своим же связям — у системы нет. Потому что у неё не связей в этом смысле, а один сплошной распределённый пласт чисел. Она не «вспоминает» так, как мы: она вычисляет. Ответ рождается из того, насколько правдоподобно сумма всех весов реагирует на входящую строку. Это не прогулка по карте города — это мгновенный срез плотности, в котором все улицы города сдавлены в одном вокзале.

И здесь корень расхождения. Модель, при всей ширине, не может перестроить себя: её «граф» запечён в матрице и меняется только внешним дообучением. Человек же перестраивает свой граф изнутри, самим фактом мышления — он сам себе и лошадь, и тренер.

Почему они оба важны

Не будем сводить всё к противостоянию. Это два разных механизма, и путать их — значит строить ложные ожидания. А вот сочетать — самое интересное.

Модель даёт человеку невообразимую ширину: она видела почти всё, что написано человечеством. Это телега, гружённая с верхом, — в ней есть всё, но сама по себе она не двинется с места. Человек даёт глубину: он живёт в своём графе и перестраивает его изнутри, но его граф, как ни крути, ограничен одной жизнью и одной картой.



Самое интересное — не гонка между ними. Самое интересное — это гибрид: когда широта модели встречается с глубиной человека, а агент выступает тем, кто запрягает. Из гружёной телеги широты агент на лету собирает плотный граф под живую задачу — и тогда она наконец становится упряжью: человеческая связность тянет, агент держит натяжение, а широта модели везёт груз, который в одиночку не сдвинуть.

Агент: временный граф в контексте

А теперь — то, ради чего эта статья вообще затевалась: агент. Потому что в агентных системах я вижу попытку — частично удачную — наложить граф поверх массива весов.

Что делает агент? У него нет врождённого графа. Но у него есть **контекст** — окошко, в которое можно что-то додать извне. И вот тут происходит трюк: агент берёт нужные куски знания, растягивает их по этому окошку и связывает там — не в весах, а прямо в окне внимания. Достал документ, прицепил к нему инструкцию, рядом положил пример, обвязал вывод. Задача решается не «за счёт весов», а «за счёт того, что в контексте собрали связный кусок».

Это и есть попытка **уплотнить связи под конкретную задачу, чтобы всё влезло в контекст**. Человек уплотняет граф, перестраивая внутренние связи. Агент уплотняет граф, перекладывая связи в открытый контекст. Разница тонкая, но решающая: человеческий граф — часть субъекта; агентский граф — временная сборка, которую собирают и разбирают в каждом окне заново.

И вот почему агент «выигрывает» в прагматике: он может каждый раз собрать самый плотный подграф под задачу. Он не обязан жить в нём — он его призывает

по вызову. Это как иметь не одну карту города, а умение за минуту распечатать точный маршрут до любого адреса, не храня атласа целиком в голове.

Самоприменение: почему мы вообще есть

Осталась одна деталь, без которой история неполна: почему живой граф вообще способен сам себя перестраивать? Ответ тянет вниз, к логике.

У живого познания нет разделения «чем думаю» и «о чём думаю» — на самом дне это одно и то же. Правило и материал, которое оно обрабатывает, сделаны из одной ткани: то, что ты понял вчера, становится тем, чем ты думаешь сегодня. Поэтому такая система может менять себя прямо на ходу: мысль выходит из неё и снова входит как материал для следующей. Ей не нужен кто-то снаружи, кто стоял бы над ней и правил правила, — она сама себе и строение, и материал, и правка.

И вот тут проходит граница. Человеческий граф — самоприменяемый: он меняет себя самим актом мышления. А модель — петля с внешней правкой: изменить её может только внешнее дообучение, а не собственный акт вывода. В этом смысле «самоусовершенствование», которое сейчас пытаются приделать нейросетям извне, — это костыль, приставленный сбоку: правку всё равно должен держать внешний механизм. А у человеческого графа этого барьера нет — он уже самоприменяемый.

Возвращаясь к графу

Мы начали с картинки про телегу и лошадь и дошли до способа, которым каждый из нас держит мир.

Настоящее «понимать» — это всегда держать мир как связный граф, в котором каждое понятие живёт на своих нитях, а нити — это прожитое. Широта без связности — это амбар, полный зерна, из которого нельзя испечь хлеб, потому что нет мельницы и нет мешка. Глубина без ширины — это мельница, которая мелко мелет лишь одно зерно, потому что другого ей не привозили.

> Понимание — не ширина, а связность. Плотность связи создаёт реальность.

И, может быть, мудрость нашего времени не в том, чтобы выбрать: стать амбаром или стать деревом. А в том, чтобы научиться быстро доращивать связи в своей голове — плотно, живо, через опыт — и не бояться впустить помощника, который умеет распечатывать точные маршруты из внешней ширины, когда задача требует скорости, а не глубины корней.

И вот тут агент, ради которого всё затевалось, наконец встаёт на своё место. Он не заменяет живой граф — у него самого его нет. Он тот, кто на лету собирает временные связки из чужой ширины и подносит их живой голове: достал по

запросу, связал под задачу, показал. Плоская широта модели через него становится упряжью — точным подграфом маршрута, когда он нужен, — а решает всё равно живой граф, который умеет вплести этот маршрут в свою карту.

Потому что здоровый порядок здесь один, и это не про то, что одно лучше другого. Это про то, что телега стоит позади лошади: связность тянет, широта везёт груз, агент держит натяжение, а живая голова решает, какую дорогу и под какую ношу открывать.

@Tottop10 · серия «Как поместить мир в голову» · свободно распространяется на условиях CC BY 4.0