УДК: 81'35:004.8 : 161.7 : 004.895
DOI: 10.15372/PS20250503
EDN: VVGZDP

П.Н. Барышников

СЕМАНТИКА ИНТЕНСИОНАЛЬНЫХ КОНТЕКСТОВ И ГЕНЕРАТИВНЫЙ ИИ[1]

В данной работе рассматривается проблема семантики интенсиональных контекстов, в особом ракурсе, связанном с принципами работы генеративного искусственного интеллекта (ИИ). Интенсиональные контексты, такие как верования, желания, знания и убеждения представляют особую сложность для современных языковых моделей, поскольку требуют учета значений выражений в логических возможных мирах или когнитивных состояниях субъектов. Данная работы – это своеобразный эскиз к постановке проблемы, у которой есть инженерное и философское измерение. Здесь анализируются механизмы работы трансформеров которые используют контекстуальные эмбеддинги для моделирования значений слов через self-attention. Выявлено, что современные языковые модели способны эффективно обрабатывать анафорические зависимости и контекстуальные связи, но при этом они сталкиваются с ограничениями при интерпретации интенсиональных конструкций. Особое внимание уделяется экспериментам с векторными представлениями объектов в многомерном пространстве, в ходе которых при различении субъективных верований и объективной реальности возникают затруднения. Характер затруднений указывает на то, что работа с интенсиональными контекстами требует не только простого анализа вероятностных связей между словами, но и более глубокого понимания семантики языковых выражений.

Ключевые слова: интенсиональный контекст, генеративаный ИИ, эмбеддинг, семантика возможных миров.

[1] Исследование выполнено за счет гранта Российского научного фонда № 24-28-00540, https://rscf.ru/project/24-28-00540/

P.N. Baryshnikov

SEMANTICS OF INTENSIONAL CONTEXTS AND GENERATIVE AI[1]

This paper examines the problem of semantics of intensional contexts from a special perspective related to the principles of operation of generative artificial intelligence (AI). Intensional contexts such as beliefs, desires, knowledge and convictions pose a particular challenge for modern language models, since they require taking into account the meanings of expressions in logical possible worlds or cognitive states of subjects. This paper is a kind of sketch for the formulation of a problem that has an engineering and philosophical dimension. Here, we analyze the mechanisms of operation of transformers that use contextual embeddings to model the meanings of words through self-attention. It was revealed that modern language models are able to effectively process anaphoric dependencies and contextual connections, but they face limitations when interpreting intensional constructions. Particular attention is paid to experiments with vector representations of objects in multidimensional space, during which difficulties arise when distinguishing between subjective beliefs and objective reality. The nature of the difficulties indicates that working with intensional contexts requires not only a simple analysis of probabilistic connections between words, but also a deeper understanding of the semantics of linguistic expressions.

Keywords: intensional context, generative AI, embedding, possible worlds semantics.

1. Источник вдохновения

Тема, вынесенная в заголовок данной работы, в силу относительной «молодости» технологий генеративного искусственного интеллекта, достаточно нова. Пока трудно обнаружить полноценного набора авторитетных источников по этой проблеме. Скорее всего такая ситуация сложилась потому, что, с одной стороны, архитекторы трансформеров не видят пользы в традиционных вопросах формальной семантики. С другой стороны, традиционные для логики и философии семантические вопросы очень уж видо-

[1] The research was supported by the grant of the Russian Science Foundation No. 24-28-00540, https://rscf.ru/project/24-28-00540/

изменяются при перенесении их в поле нейросетевых технологий. [1] Модели на основе трансформеров способны частично улавливать интенсиональные контексты через контекстуализированные эмбеддинги, под которыми понимается числовое векторное представление информационных объектов. Но совместимость формальных аппаратов недостаточно проанализирована. Представленный текст можно рассматривать как эскиз к более широкому исследованию.

Идея написать тезисы по проблеме интенсиональных контекстов в современных методах NLP (Natural Language Processing), в частности при использовании генеративных трансформеров, возникла после прочтения обзорной работы П. Куслия и Е.Вострико-вой [2]. В указанной работе представлена широкая экспозиция проблемы интенсиональных контекстов, с выделением механизмов семантических ограничений, определяющих совместимость различных классов интенсиональных глаголов с декларативными и вопросительными придаточными предложениями. Мы видим, что традиционные синтаксические объяснения несовместимости некоторых глаголов с вложенными вопросами являются недостаточными, поскольку грамматическая корректность таких конструкций определяется их семантикой, включая пресуппозиции, модальность и функциональные особенности отрицания. Важным и эффективным теоретическим инструментом является концепция L-аналитичности. Основная идея этого подхода состоит в том, что некоторые выражения остаются грамматически некорректными не из-за их синтаксической структуры, а потому, что они семантически тривиальны (либо тавтологичны, либо противоречивы) в силу особенностей их функциональных элементов. Это ограничение исходит из свойств логической структуры предложения, а не его лексического наполнения. Концепция L-аналитичности близка к идеям Куайна и его критике аналитичности, [4] но фокусируется на формальных ограничениях грамматики. Например, в некоторых работах показывается, что естественный язык исключает не просто аналитические истины (как в логическом позитивизме), а L-аналитичные структуры, которые неизбежно приводят к бессмысленности. [3] Из этого можно заключить, что грамматическая корректность интенсио-

нальных конструкций не сводится к синтаксическим ограничениям, а определяется более глубокими семантическими механизмами, связанными с пропозициональными установками, пресуппозициями и взаимодействием отрицания с кванторными и модальными элементами.

Интенсиональные контексты представляют собой нетривиальную задачу для генеративного искусственного интеллекта (ИИ), поскольку они требуют учета значений выражений в разных возможных мирах или контекстах, а не только в одном фиксированном значении. Эта инженерная задача позволяет сформулировать философскую проблему:

Какой формальный механизм позволит генеративному искусственному интеллекту адекватно моделировать семантику интенсиональных контекстов, учитывая их зависимость от возможных миров и когнитивных состояний субъектов?

Для обеспечения корректной обработки пропозициональных установок, модальных конструкций и интенсиональных выражений в условиях неопределенности контекста эта проблема требует объединения моделей представления знания (например, логики возможных миров, семантических графов, вероятностных моделей) с принципами работы генеративных моделей. Какие есть варианты решения? Рассмотрим некоторые принципы анализа контекстуального окружения современными LLM.

2. Моделирование контекста

ИИ может использовать контекстуальные подсказки для уточнения значений выражений. В LLM (Large Language Models – Большие языковые модели) слова интерпретируются на основе их окружения в тексте, что позволяет моделировать различие между буквальным значением и значением в определенном контексте (например, в гипотетическом или возможном мире). При анализе интенсиональных контекстов, таких как «думать», «верить» или «желать», модель ориентируется на слова, указывающие на ментальные состояния или гипотетические ситуации. При этом языковые модели не имеют встроенного понятия о «возможных мирах». Они могут только подстраиваться под вероятностные связи в структурах данных.

Перед тем, как переходить к примерам со сложными интенсиональными контекстами, рассмотрим два примера с анафорическими зависимостями:

1. *Чем больше войско у генерала, тем* ***ему труднее****.*

2. *Генералы убеждены, что чем больше в войске солдат, тем* ***им труднее****.*

Современные NLP-модели (BERT, GPT, T5) работают с self-attention, определяя вероятностную связь между словами. Self-attention представляет собой механизм в архитектурах нейронных сетей, который позволяет модели определять и учитывать взаимосвязи между различными частями входного текста. Такой механизм позволяет каждому слову (или токену) обращаться ко всем другим словам в предложении, вычисляя их «влияние» или «важность» для понимания данного слова[1]. Формула для трансформеров выглядит так:

$$Attention\,(Q,K,V) = softmax\left(\frac{QK^t}{\sqrt{d^k}}\right)V$$

где:

- Q(Query), K(Key), V(Value) – векторы слов, вычисленные на основе контекстного представления.
- d_k – размерность эмбеддингов (используется для масштабирования).

BERT обучается на предсказании замаскированных слов и следующего предложения. Важно понимать, как модель определяет референт местоимения «ему»? В первую очередь создаются векторные представления слов (Word Embeddings):

Таблица 1

Пример (1)	**Пример (2)**
• "генерал" → v_g • "войско" → v_w • "ему" → v_p	• "генералы" → v_g • "солдаты" → v_w • "им" → v_p

[1] Отметим, что некоторые модели, например, BERT используют маскированное языковое моделирование (MLM). В отличие от традиционных языковых моделей (GPT), которые обычно прогнозируют следующее слово в последовательности, маскированное языковое моделирование позволяет обучать модель на основе всей доступной информации из контекста, а не только из прошлого. Из-за этого результаты анализа в разных моделях могут значительно расходиться. Но для данного этапа постановки проблемы это имеет значение, которым можно пренебречь.

Далее определяется вероятностная связь через self-attention: для слова «ему» модель оценивает связь с «генералом» и «войском». Вероятность связи вычисляется как $Q_p \cdot K_g$ для генерала) и $Q_p \cdot K_w$ (для войска). Далее подсчитывается связь через Attention Score, и на основании этого делается вывод.

Таблица 2.

Пример	**Слово**	**Связь с “ему” (Attention Score)**
(1)	генерал	0,85
(1)	войско	0,15
Пример	**Слово**	**Связь с “им” (Attention Score)**
(2)	генералы	0,23
(2)	солдаты	0,77

Сразу отметим, что результаты подсчёта «Внимания» на разных моделях могут быть разными. Приведённые выше примеры – результат обработки запросов на ChatGPT4o. Недавно вышедшая китайская модель Qwen2.5-Plus в обоих примерах присвоила словам «генерал» и «генералы» индекс «Счёта внимания» 0,6 относительно слов «ему» и «им» соответственно. Большую роль играют данные, на которых обучалась модель. Вероятностные расстояния между словами, рассчитываются именно на основе зависимостей в обучающем наборе данных. Собственно, существует несколько аспектов, которые влияют на интерпретацию LLM предложений типа (2). Модель распознаёт:

1. **Синтаксический аспект:** «Генералы убеждены» создает интенсиональный контекст, в котором мнение (убежденность) принадлежит генералам.
2. **Семантический аспект:** В некоторых контекстах можно интерпретировать, что «им труднее» относится к генералам, так как они несут больше ответственности при увеличении численности войска.
3. **Аспект обучения:** Если обучающие данные содержали больше примеров, где дательный падеж «им» чаще соотносится с подлежащим («генералы»), модель будет склонна делать аналогичное предсказание.

4. **Информационно-дистрибутивный аспект**: В обучающем корпусе модель, возможно, чаще встречала конструкции, где «генералы» и «им» тесно связаны.

5. **Аспект референциальной близости**: Трансформеры иногда предпочитают связывать местоимения с ближайшим возможным референтом в рамках одной смысловой структуры.

3. Возможные миры в многомерном векторном пространстве[1]

Возьмем более сложный пример:

(1) *Алиса верит, что единорог живет в ее саду, но не знает, что это тот же самый единорог, о котором говорит Боб.*

В стандартной экстенсиональной логике, если два языковых выражениях указывают н один и тот же объект, то они взаимозаменяемы без потери смысла всего высказывания. В высказывании (3) утверждается, что единорог в саду Алисы является, тем же объектом, о котором говорит Боб, но Алиса об этом не знает. ИИ, выстраивая векторные предсказания последовательностей токенов, может ошибочно предположить, что Алиса знает, что объем понятия «единорог, живущий в саду Алисы» и «единорог, котором горит Боб» равнозначны. Генеративный ИИ, обученный на корпусах текстов, может не учитывать ограниченную перспективу субъекта**,** что может привести к ошибочному выводу. Если ИИ использует векторное представление значений, то «единорог Алисы» и «единорог Боба» могут быть слишком близки в многомерном пространстве, что приведет к ложному объединению понятий.

Проведем эксперимент. У нас есть два объекта: «Единорог Алисы» – объект, который согласно вере Алисы, живёт в её саду, и «Единорог Боба» – объект, о котором говорит Боб. Каждый из этих объектов может быть представлен как вектор в многомерном пространстве, основанный на контексте их упоминания. Для большей точности мы будем проводить измерение косинусной близости между «единорогом Алисы» и «единорогом Боба» на предобученой англоязычной версии модели BERT на примере:

[1] Эксперименты проводились на модели https://chat.qwenlm.ai/ [5]

(2) *Alice believes that a unicorn lives in her garden, but she does not know that it is the same unicorn that Bob talks about.*[1]

Код позволяет вычислить семантическое расстояние между двумя упоминаниями слова «unicorn» в предложении с использованием BERT. Предложение токенизируется, преобразуется в векторы с помощью модели, и для каждого «unicorn» извлекаются соответствующие векторы. Затем через косинусную близость оценивается их сходство, что позволяет количественно проанализировать семантическую связь этих упоминаний в интенсиональном контексте.

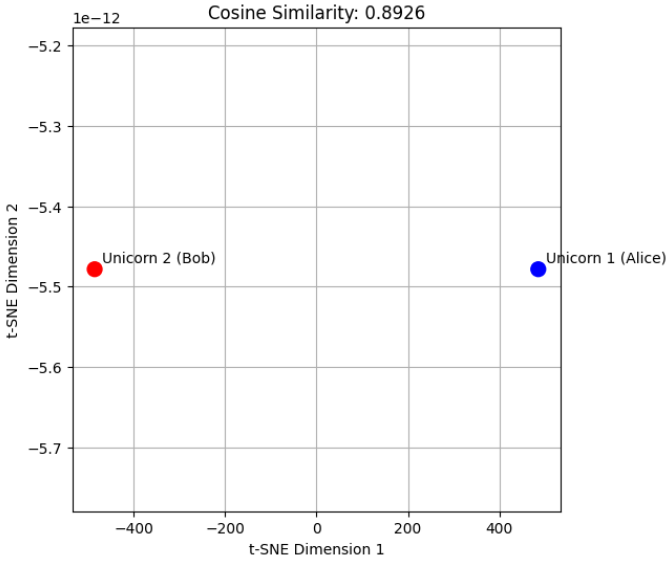

Рис. 1. Косинусная близость Unicorn 1(Alice) и Unicorn 2 (Bob)

Косинусная близость ≈ 089: Векторы «единорог Алисы» и «единорог Боба» в англоязычной версии достаточно близки семантически, что может отражать их потенциальную идентичность. Однако, для языковой модели логические интенсиональные ограничения остаются недоступными. Без явной аннотации

[1] https://colab.research.google.com/drive/15LNhUs89YDQxhn5C1-GEPwoh5R2Y5ZkW#scrollTo=euh8psa0QxF9&line=42&uniqifier=1

или заданной онтологии модели не могут предсказать, что интенсиональный контекст потребует учета восприятия или убеждений субъекта, а не объективного состояния дел.

4. Убеждения Алисы и идентичность объектов

Для формализации примера (3) определим термины:

- w: Возможный мир.
- E(w): Единорог в мире w.
- G(x): x живет в саду Алисы.
- B(x): x – единорог, о котором говорит Боб.
- $K_A(\phi)$: Алиса знает, что ϕ истинно.
- $B_A(\phi)$: Алиса верит, что ϕ истинно.

Тогда предложение можно разбить на две части:

1. BA ($\Box x[E(x) \Box G(x)]$) – Алиса верит, что существует единорог x, который живет в её саду.
2. $\neg$ KA($(\Box x \Box y[(E(x) \Box G(x)) \leftrightarrow (E(y) \Box B(y))]$) Алиса не знает, что единорог в её саду идентичен единорогу, о котором говорит Боб.

При такой формализации выявляется три проблемы. Во-первых, интенсиональность веры подразумевает возможность несоответствия между верованиями субъекта и объективной реальностью. Во-вторых, определение идентичности объектов в интенсиональных контекстах требует специальной семантики, поскольку субъект может не осознавать идентичность двух объектов даже при их фактическом совпадении. В-третьих, различие между знанием и верой заключается в том, что знание предполагает уверенность в истинности утверждения, тогда как вера может быть ошибочной. Таким образом, Алиса может верить в существование единорога, но не иметь достоверного знания о его связи с единорогом Боба.

Как мы показали в эксперименте, LLM сталкиваются с трудностями при установлении идентичности объектов и при анализе анафорических зависимостей, поскольку они не имеют доступа к формальной логике возможных миров или другим механизмам контекстуального анализа. Вместо этого модели полагаются на контекстуальные эмбеддинги, которые могут быть недостаточно точными для решения таких задач. Отметим, что запрос на созда-

ние эмпатичного и контекстуально чувствительного ИИ существует давно. [6] Способность извлекать смысл из контекста вместо простого распознавания паттернов данных позволило бы системам быть более гибкими и адаптивными. Распознавание представлений, намерений, убеждений и прочих ментальных состояний в ходе принятия решений также открыло бы новые возможности для ИИ, улучшая взаимодействие между человеком и машиной.

Итак, Большие языковые модели основываются на статистических закономерностях, извлеченных из больших корпусов текста. Они генерируют ответы, опираясь на вероятностные связи между словами, но не обладают истинным пониманием содержания. Это становится критичным в случае интенсиональных контекстов, где значение предложения зависит не только от составляющих его слов, но и от ментальных состояний субъекта (например, знания, веры, желания). Такие утверждения как (3) и (4) требуют интерпретации веры Алисы как модального оператора, который может не соответствовать положению дел в реальном мире. LLM могут воспроизводить подобные конструкции, однако их способность различать истинное знание и ошибочную веру остается ограниченной.

Литература

1. *Havlík V.* Meaning and understanding in large language models. // Synthese, 2024. Vol. 205, No. 1. P. 9-29.

2. *Куслий П.С., Вострикова Е.В.* Семантика интенсиональных контекстов: современные проблемы и дискуссии. // Философия науки и семантика. / Научн. ред. и сост. Р.Э. Бараш, Е.В. Вострикова, П.С. Куслий. М.: Русское общество истории и философии науки – 2020. С. 292-327.

3. *Куайн У.В.О.* С точки зрения логики. 9 логико-философских очерков. – Томск: Изд-во Томского университета. 2003.

4. *Gajewski J.* On analyticity in natural language. 2002. Citation Key: Gajewski2004OnAI. https://api.semanticscholar.org/CorpusID:85508868

5. *Team Q.* Qwen2.5 technical report. 2024. arXiv preprint arXiv:2412.15115. https://arxiv.org/abs/2412.15115

6. *Bennett M.T., Maruyama Y.* Intensional Artificial Intelligence: From Symbol Emergence to Explainable and Empathetic AI. // Intensional Artificial Intelligence. 2021.: arXiv. https://arxiv.org/abs/2104.11573

References

1. Havlík V. (2024) *Meaning and understanding in large language models*. Synthese, Vol. 205, No. 1, P. 9.

2. Kusliy P.S., Vostrikova E.V. (2020) *Semantics of intensional contexts: modern problems and discussions*. Philosophy of science and semantics: monograph / Scientific. ed. and comp. R.E. Barash, E.V. Vostrikova, P.S. Kusliy. – : "Truth. Science. Reason". – M .: Russian Society for the History and Philosophy of Science. – pp. 292-327. (In Russ.)

3. Gajewski J. (2004) *On analyticity in natural language*. Citation Key: Gajewski2004OnAI. https://api.semanticscholar.org/CorpusID:85508868

4. Quine W.V.O. (2003) From the point of view of logic. 9 logical-philosophical essays: Library of analytical philosophy. – Tomsk: Tomsk University Press. – 166 p. (In Russ.)

5. Team Q. (2024) *Qwen2.5 technical repor*t. arXiv preprint arXiv:2412.15115. https://arxiv.org/abs/2412.15115

6. Bennett M.T., Maruyama Y. (2021) *Intensional Artificial Intelligence: From Symbol Emergence to Explainable and Empathetic AI*. Intensional Artificial Intelligence. – arXiv. https://arxiv.org/abs/2104.11573

Информация об авторе

Барышников Павел Николаевич – д. филос. н., доцент, профессор кафедры исторических, социально-философских дисциплин, востоковедения и теологии Пятигорского государственного университета.

pnbaryshnikov@pgu.ru

Information about the authors

Baryshnikov Pavel N. – Doctor of Philosophy, Associate Professor, Professor of the Department of Historical, Social and Philosophical Disciplines, Oriental Studies and Theology of Pyatigorsk State University.

pnbaryshnikov@pgu.ru

Дата поступления 12.03.2025
Принята к печати 11.12.2025