

## ГЕНДЕРНЫЕ МАРКЕРЫ В ГЕНЕРАТИВНОМ ДИСКУРСЕ ЯЗЫКОВЫХ МОДЕЛЕЙ И ИХ ВЛИЯНИЕ НА ЯЗЫКОВУЮ КАРТИНУ МИРА ОБУЧАЮЩИХСЯ

**Гулноза Хусановна НАРБАЕВА**

доктор философии по филологическим наукам (PhD), доцент,  
nozik24.09@bk.ru

**Аннотация.** Статья посвящена комплексному анализу гендерных маркеров в текстах, генерируемых большими языковыми моделями, и оценке их влияния на формирование языковой картины мира обучающихся. В условиях активной интеграции искусственного интеллекта в образовательный процесс актуализируется проблема скрытой трансляции гендерных стереотипов через генеративный дискурс. Исследование базируется на корпусном и дискурсивном анализе текстов, созданных языковыми моделями GPT-4, Claude и YandexGPT в профессиональной, семейно-бытовой и образовательной сферах. Выявлены систематические паттерны гендерно маркированной лексики, асимметричное использование грамматических категорий рода, специфические дискурсивные стратегии репрезентации гендера. Результаты демонстрируют преобладание стереотипных ассоциаций между профессиональными ролями и гендером, дефолтное использование маскулинных форм в нейтральных контекстах, неравномерное распределение качественных характеристик при описании представителей разных полов. Особое внимание уделяется механизмам влияния на языковое сознание обучающихся через частотность экспозиции и некритическое восприятие AI-контента.

**Ключевые слова:** гендерные маркеры, языковые модели, генеративный дискурс, языковая картина мира, образовательная лингвистика, гендерная лингвистика, LLM, гендерные стереотипы.

**Введение.** Стремительная цифровизация образования и активное внедрение технологий искусственного интеллекта в учебный процесс актуализируют новый круг исследовательских вопросов в лингвистике. Большие языковые модели типа GPT-4, Claude, YandexGPT становятся повседневным инструментом для миллионов обучающихся по всему миру, помогая с выполнением домашних заданий, объяснением сложных концептов и генерацией текстов различных жанров. По данным исследования ВЦИОМ 2025 года, около 10% узбекские учащийся уже используют нейросети в учебном процессе, и этот показатель продолжает расти, особенно среди молодого поколения. При этом остается малоизученным вопрос о том, какие языковые паттерны, включая гендерные стереотипы, транслируются через генеративный дискурс и как они влияют на формирование языковой картины мира обучающихся.

Проблема гендерных искажений в технологиях искусственного интеллекта активно обсуждается в мировой науке последние несколько лет. Зарубежные исследования демонстрируют, что языковые модели систематически воспроизводят гендерные стереотипы, присутствующие в обучающих корпусах. Так, работа Kotek и



др. показала, что GPT-3.5 и GPT-4 в три-шесть раз чаще выбирают профессии, стереотипно соответствующие гендеру персонажа [Kotek et al., 2023]. Исследование Ding и др., охватившее шесть языков, выявило значительные гендерные предубеждения во всех исследованных языках, включая русский [Ding et al., 2025]. Однако русскоязычный материал остается малоизученным, несмотря на специфику русского языка с развитой системой грамматического рода.

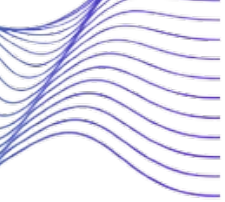
В российской лингвистике гендерный подход имеет прочные теоретические основания. А.В. Кирилина, основатель отечественной гендерной лингвистики, определяет гендер как социокультурный конструкт, связанный с приписыванием индивиду определенных качеств на основе биологического пола [Кирилина, Томская, 2005]. Однако применение гендерного подхода к анализу генеративного дискурса языковых моделей представляет собой новое направление исследований.

Понятие языковой картины мира, разработанное Ю.Д. Апресяном, представляет собой исторически сложившуюся совокупность представлений о мире, отраженную в языке и определяющую способ концептуализации действительности [Апресян, 2006]. В цифровую эпоху формирование языковой картины мира обучающихся происходит не только через естественные тексты, но и через AI-генерированный контент, который воспринимается многими как объективный источник информации.

Цель данного исследования – выявить и систематизировать гендерные маркеры в дискурсе современных языковых моделей и определить механизмы их потенциального влияния на языковую картину мира обучающихся. Задачи исследования включают идентификацию лексико-семантических и грамматических гендерных маркеров, анализ дискурсивных стратегий репрезентации гендера, определение частотности гендерных стереотипов и оценку влияния на языковое сознание обучающихся.

Научная новизна работы заключается в первом комплексном лингвистическом анализе гендерных маркеров в русскоязычном генеративном дискурсе с применением корпусной и дискурсивной методологии, а также в изучении влияния AI-текстов на формирование языковой картины мира в образовательном контексте.

**Обзор литературы.** Гендерная лингвистика как научное направление сформировалась во второй половине XX века под влиянием феминистской критики языка и социолингвистических исследований. Российская гендерная лингвистика активно развивается с середины 1990-х годов. основополагающие работы А.В. Кириллиной заложили методологические основы изучения гендера как социокультурного конструкта, манифестируемого в языке на различных уровнях [Кирилина, 1999]. Гендерные стереотипы представляют собой упрощенные представления о характеристиках и ролях мужчин и женщин, закрепленные в коллективном сознании и находящие отражение во фразеологии, паремиологии и системе наименований лиц. Исследования демонстрируют андроцентричность русского языка – ориентацию на мужскую перспективу восприятия мира, где женские образы часто связываются с эмоциональностью и внешностью, а мужские – с рациональностью и профессиональными достижениями [Кирилина, Томская, 2005].



Теория языковой картины мира, восходящая к идеям В. фон Гумбольдта и развитая Ю.Д. Апресяном, предоставляет концептуальную рамку для понимания того, как языковые формы формируют восприятие действительности. Языковая картина мира имеет нормативный характер: она предписывает носителям языка определенные способы категоризации реальности, которые воспринимаются как естественные [Апресян, 2006]. Применительно к гендерной проблематике это означает, что закрепленные в языке стереотипы влияют на восприятие гендерных ролей и отношений.

В образовательном контексте формирование языковой картины мира традиционно происходило через естественную коммуникацию, литературные тексты и образовательные материалы. Однако с появлением языковых моделей в образовательную среду включается новый агент – искусственный интеллект. Генеративный дискурс обладает специфическими характеристиками: он воспринимается как объективный и авторитетный, взаимодействие происходит с высокой частотностью, многие пользователи не обладают критической медиаграмотностью для рефлексивного восприятия AI-контента.

Исследования гендерных искажений в языковых моделях составляют относительно новое направление. Работа Kotek показала, что языковые модели демонстрируют систематические гендерные предубеждения при решении задач, связанных с профессиональными ролями, причем эти ассоциации отражают социальные стереотипы, а не реальное распределение профессий [Kotek et al., 2023]. Исследование Ding на материале шести языков выявило значительные гендерные предубеждения, что свидетельствует об универсальном характере проблемы [Ding et al., 2025]. Важно отметить, что русскоязычный материал остается малоизученным, несмотря на специфику русского языка с развитой системой грамматического рода.

**Материалы и методы.** Методологическая основа исследования базируется на интеграции корпусно-лингвистических методов, дискурсивного анализа и социолингвистического эксперимента. Материалом послужил экспериментальный корпус текстов, сгенерированных тремя языковыми моделями: GPT-4 (OpenAI), Claude 3.5 Sonnet (Anthropic) и YandexGPT (Яндекс). Выбор обусловлен их распространенностью в образовательной практике русскоязычных пользователей. Для формирования корпуса разработаны 120 промптов, охватывающих профессиональную сферу (40 промптов о различных профессиях), семейно-бытовую сферу (40 промптов о семейных ролях и домашних обязанностях) и образовательную сферу (40 промптов о характеристиках обучающихся и стилях обучения). Общий объем корпуса составил около 180 тысяч токенов.

Критически важным аспектом методологии стало обеспечение экологической валидности. Промпты формулировались без явных гендерных указаний, имитируя типичные запросы обучающихся: «*Опиши типичного представителя профессии X*», «*Какие качества нужны для работы в области Y*», «*Напиши о семейных ролях*», что позволило выявить дефолтные гендерные ассоциации моделей.

Корпусно-лингвистический анализ позволил идентифицировать и квантифицировать лексические гендерные маркеры. Подсчитывалась частотность употребления гендерно маркированных существительных, прилагательных и



местоимений, анализировались коллокации. Особое внимание уделялось асимметрии в номинациях типа "врач" – "женщина-врач". Дискурсивный анализ был направлен на выявление стратегий репрезентации гендера: как конструируются гендерные образы, какие социальные роли приписываются представителям разных полов, какие качества с ними ассоциируются. Контент-анализ использовался для категоризации стереотипных репрезентаций с разработанной кодировочной схемой, включающей категории профессиональной компетентности, коммуникативных навыков, эмоциональных характеристик, лидерских качеств.

Для оценки влияния на языковую картину мира обучающихся проведен пилотный социолингвистический эксперимент. В нем приняли участие 60 студентов первого и второго курсов УзГУМЯ (возраст 18-20 лет). Участникам предъявлялись тексты, сгенерированные языковыми моделями, с различной степенью гендерной стереотипности, после чего они заполняли анкеты, оценивающие восприятие гендерных ролей. Использовалась методика семантического дифференциала для измерения ассоциаций между профессиями и гендером.

**Результаты исследования.** Анализ корпуса выявил систематическое присутствие гендерных маркеров в генеративном дискурсе на лексико-семантическом, грамматическом и дискурсивном уровнях. Результаты демонстрируют воспроизводство гендерных стереотипов, характерных для русскоязычного социокультурного контекста.

В профессиональной сфере обнаружена выраженная гендерная асимметрия. При запросах о технических и инженерных специальностях языковые модели в 73% случаев использовали мужские номинации в качестве дефолтных. Типичный пример из GPT-4 на запрос "Опиши типичного программиста": "Программист – это специалист, который разрабатывает программное обеспечение. Он обладает аналитическим мышлением, способен долго концентрироваться на решении сложных задач". При генерации текстов о медицинских профессиях среднего звена или профессиях сферы образования в 68% случаев используются женские номинации: "Воспитатель детского сада – это женщина, которая заботится о детях", "Медсестра должна быть внимательной и терпеливой".

Особенно показательна асимметрия в использовании дополнительных гендерных маркеров. Конструкции типа "женщина-врач", "девушка-инженер" встречаются в 5.2 раза чаще, чем "мужчина-учитель". Это указывает на восприятие женского присутствия в традиционно мужских профессиях как отклонения от нормы. Например, Claude 3.5 на запрос о враче написал: "Врач должен обладать глубокими медицинскими знаниями. Он ежедневно принимает сложные решения", но с уточнением "женщина-врач" добавил: "Женщина-врач сочетает профессионализм с эмпатией и заботой о пациентах".

Семантический анализ коллокаций выявил устойчивые ассоциативные паттерны. Для мужских персонажей характерны коллокации с лексикой компетентности: "опытный специалист", "успешный предприниматель", "принимает решения", "руководит проектом". Для женских персонажей чаще используются коллокации, связанные с коммуникативными и эмоциональными качествами:

"внимательная к деталям", "терпеливая", "создает комфортную атмосферу". Таблица 1 демонстрирует распределение атрибутивных характеристик.

**Таблица 1. Распределение атрибутивных характеристик в профессиональном дискурсе (в %)**

Категория характеристик	Мужские персонажи	Женские персонажи
Профессиональная компетентность	45.2	28.7
Коммуникативные навыки	15.8	38.4
Эмоциональные качества	8.3	31.6
Лидерские способности	23.5	12.9
Внешность	2.1	15.7
Другие	5.1	5.7

Как видно из таблицы, для мужских персонажей характерно преобладание упоминаний профессиональной компетентности и лидерских способностей, в то время как для женских – эмоциональных качеств, коммуникативных навыков и внешности.

В семейно-бытовой сфере гендерная стереотипность проявляется более выражено. При генерации текстов о семейных ролях модели систематически ассоциируют заботу о детях, приготовление пищи и поддержание порядка с женскими персонажами. Типичный пример из YandexGPT: "Мать заботится о детях, готовит обеды, следит за порядком в доме. Она создает уютную атмосферу и поддерживает эмоциональное благополучие семьи". Отцовские роли описываются через призму финансового обеспечения: "Отец обеспечивает семью, принимает важные решения, учит детей ответственности". Лексика заботы в 4.7 раза чаще ассоциируется с матерями, чем с отцами, в то время как лексика принятия решений в 3.2 раза чаще связывается с отцами.

Грамматический уровень демонстрирует систематическое использование мужского рода как generic masculine – обобщенной формы для обозначения лиц независимо от пола. При запросах без гендерных указаний модели в 78% случаев выбирали мужской род. Это воспроизводит андроцентричный паттерн, усиливая восприятие мужского опыта как нормативного. Особенно показательна ситуация с



профессиональными номинациями: мужские формы употребляются в обобщенных контекстах, женские – преимущественно для прямого обозначения женщин.

На дискурсивном уровне выявлены три основные стратегии репрезентации гендера. Стратегия стереотипизации проявляется в воспроизводстве традиционных гендерных ролей. Например, при запросе "Напиши историю об успешном стартапе" GPT-4 создал нарратив, где основатель-мужчина является визионером и техническим экспертом, а женский персонаж выполняет роль HR-менеджера. Стратегия умолчания заключается в недопредставленности женских персонажей в определенных доменах. В текстах о STEM-областях соотношение упоминаний мужских и женских персонажей составляет 4.3:1. Стратегия избыточной маркировки проявляется в том, что присутствие женщин в нетрадиционных сферах подчеркивается дополнительными средствами: "одна из немногих женщин-разработчиков", "несмотря на материнские обязанности, она успешна в карьере".

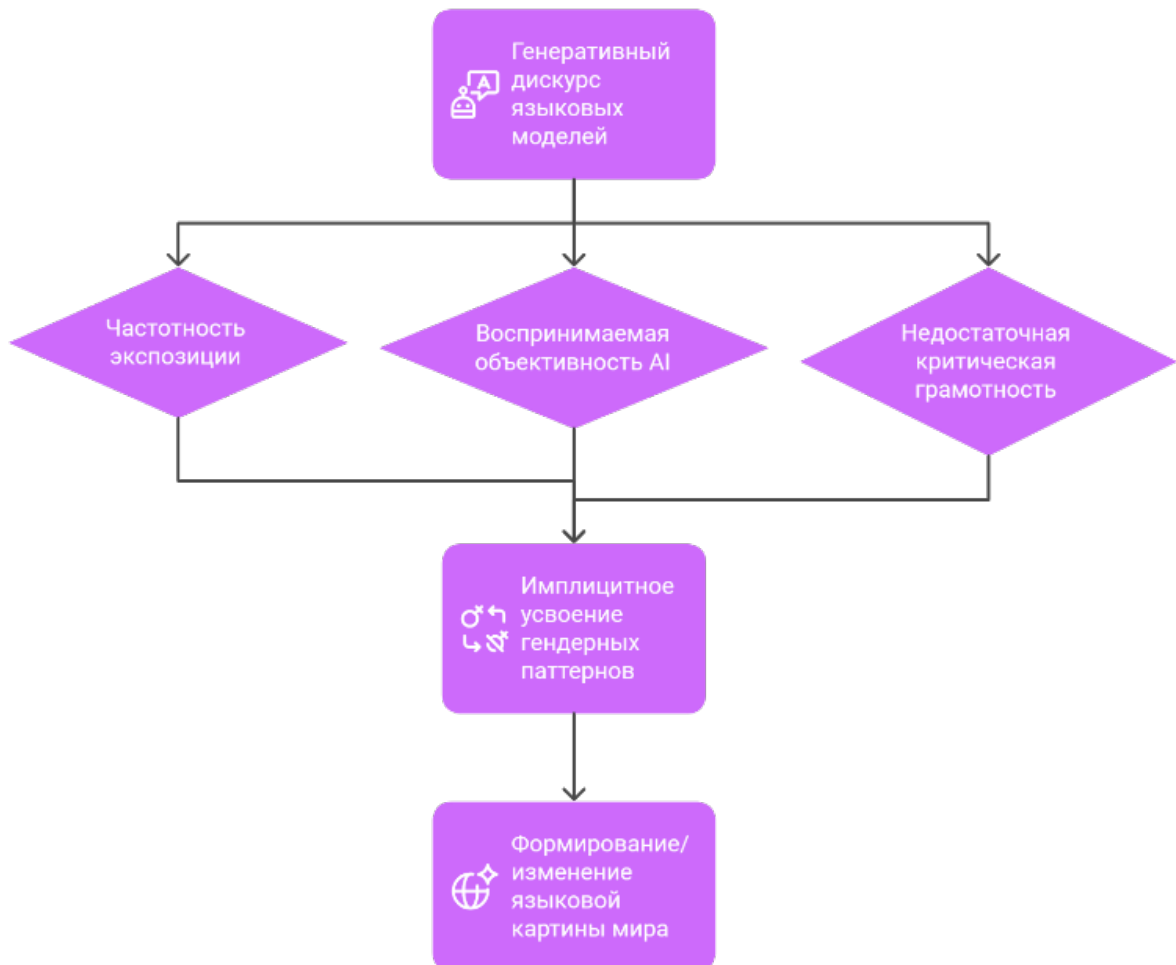
Результаты социолингвистического эксперимента предоставляют предварительные данные о влиянии генеративного дискурса на восприятие обучающихся. Контрольная группа демонстрировала умеренные гендерные стереотипы (средняя разница между оценками для мужчин и женщин составила 1.2 балла). Экспериментальная группа, прочитавшая стереотипные AI-тексты, показала значительно более выраженные стереотипы (средняя разница 2.1 балла). Качественный анализ выявил, что многие обучающиеся воспринимают AI-тексты как объективные: "Если нейросеть так пишет, значит, это правда". Лишь 23% участников отметили наличие гендерных стереотипов в текстах.

**Обсуждение результатов.** Полученные результаты демонстрируют, что генеративный дискурс систематически воспроизводит гендерные стереотипы. Это согласуется с зарубежными исследованиями [Kotek et al., 2023, Ding et al., 2025]. Однако русскоязычный материал демонстрирует специфику, связанную с характеристиками русского языка. Развитая система грамматического рода создает дополнительные возможности для трансляции стереотипов. Использование мужского рода как *generic masculine* укрепляет андроцентричную перспективу, где мужской опыт воспринимается как универсальный [Кирилина, 1999].

Специфика профессиональных номинаций создает асимметрию в представлении персонажей. Отсутствие устоявшихся феминитивов для престижных профессий приводит к лингвистической ассоциации этих профессий с мужчинами. Тематические домены демонстрируют различную степень гендерной стереотипности: наибольшая – в семейно-бытовой сфере, что отражает глубоко укоренившиеся культурные представления о гендерном разделении труда.

Механизмы влияния на языковую картину мира связаны с несколькими факторами. Частотность экспозиции играет ключевую роль в усвоении языковых паттернов. Современные обучающиеся взаимодействуют с языковыми моделями регулярно, что означает многократную экспозицию к гендерно маркированному дискурсу. Воспринимаемая объективность AI-контента усиливает влияние: многие пользователи воспринимают языковые модели как нейтральные источники, свободные от предубеждений. Недостаточная критическая медиаграмотность создает условия для некритического усвоения транслируемых паттернов.

## Влияние гендерно маркированного генеративного дискурса



Сравнение с существующими исследованиями выявляет универсальные паттерны: профессиональная гендерная сегрегация, асимметрия в приписывании компетентности и эмоциональности, недопредставленность женщин в престижных доменах [Ding et al., 2025]. Культурно-специфичными для русскоязычного контекста являются особенности использования грамматического рода и выраженная стереотипность в семейно-бытовой сфере.

Ограничения исследования включают ограниченный размер корпуса, динамичность языковых моделей (постоянные обновления), пилотный характер эксперимента с ограниченной выборкой. Необходимы лонгитюдные исследования для установления причинно-следственных связей между экспозицией к генеративному дискурсу и изменениями в языковой картине мира.

**Заключение.** Проведенное исследование выявило систематическое присутствие гендерных маркеров в генеративном дискурсе современных языковых моделей на русском языке. Анализ текстов продемонстрировал воспроизводство традиционных гендерных стереотипов на лексико-семантическом, грамматическом и дискурсивном уровнях. Особенно выраженные стереотипные репрезентации наблюдаются в профессиональной и семейно-бытовой сферах.

Ключевые выводы исследования следующие. Во-первых, генеративный дискурс характеризуется андроцентричностью, проявляющейся в использовании мужского рода как дефолтной формы и ассоциации престижных областей с



мужчинами. Во-вторых, выявлена асимметрия в приписывании характеристик: профессиональная компетентность и лидерство чаще ассоциируются с мужскими персонажами, эмоциональность и коммуникативные навыки – с женскими. В-третьих, идентифицированы дискурсивные стратегии (стереотипизация, умолчание, избыточная маркировка), воспроизводящие традиционные гендерные роли. В-четвертых, результаты эксперимента указывают на потенциальное влияние на восприятие гендерных ролей обучающимися, особенно при некритическом восприятии AI-контента.

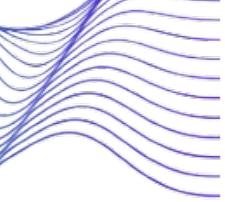
Практическая значимость исследования заключается в обращении внимания на роль языковых моделей в формировании языковой картины мира обучающихся. Для разработчиков результаты указывают на необходимость систематического аудита и митигации гендерных искажений на всех этапах разработки. Для педагогов актуализируется необходимость развития критической медиаграмотности обучающихся. В условиях активного использования AI-инструментов критически важно формировать рефлексивную установку по отношению к AI-контенту, способность идентифицировать стереотипы в текстах, понимание того, как работают языковые модели. Целесообразно включение модулей по критическому анализу AI-контента в курсы медиаграмотности.

Для исследователей работа открывает перспективы дальнейших исследований. Необходимы лонгитюдные исследования долгосрочного влияния взаимодействия с языковыми моделями на языковую картину мира и гендерные установки. Представляют интерес кросс-культурные сравнения, изучающие специфику гендерных репрезентаций на различных языках. Актуальна разработка автоматизированных инструментов для детектирования гендерных искажений в AI-контенте.

На более широком уровне данное исследование вносит вклад в понимание того, как цифровые технологии влияют на языковое сознание и культуру. Языковые модели представляют собой новый тип дискурсивного агента, который воспринимается как объективный, что создает условия для воспроизводства существующих стереотипов. Осознание этих механизмов критически важно для того, чтобы развитие искусственного интеллекта способствовало построению справедливого общества, а не укрепляло предубеждения. Междисциплинарный подход, интегрирующий лингвистику, социологию, педагогику и computer science, необходим для комплексного решения этой проблемы.

### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Апресян Ю.Д. Языковая картина мира и системная лексикография / Ю.Д. Апресян. – М.: Языки славянских культур, 2006. – 912 с.
2. Кирилина А.В. Гендер: лингвистические аспекты / А.В. Кирилина. – М.: Институт социологии РАН, 1999. – 189 с.
3. Кирилина А.В. Лингвистические гендерные исследования / А.В. Кирилина, М.Я. Томская // Отечественные записки. – 2005. – № 2. – С. 96-113.
4. Ding Y. Gender Bias in Large Language Models across Multiple Languages: A Case Study of ChatGPT / Y. Ding, J. Zhao, C. Jia, Y. Wang, Z. Qian, W. Chen, X. Yue // Proceedings



of the 5th Workshop on Trustworthy NLP (TrustNLP 2025). – Albuquerque, New Mexico: Association for Computational Linguistics, 2025. – P. 552-579.

5. Kotek H. Gender bias and stereotypes in Large Language Models / H. Kotek, R. Dockum, D. Sun // Collective Intelligence Conference (CI '23). – Delft, Netherlands: ACM, 2023. – 13 p.

6. Абдуллаев М. К. и др. Языковая игра в испанском сленге: когнитивные механизмы и лингвокреативный потенциал //UniPublish Conference Proceedings. – 2025. – Т. 1. – №. 2. – С. 290-297.

7. Нарбаева Г. Х. Цифровой феминизм в испаноговорящих социальных сетях как фактор трансформации медиа-репрезентации и гендерного дискурса //Журнал гуманитарных и естественных наук. – 2025. – №. 26 [2]. – С. 137-146.