



Большие языковые модели и текстовый анализ в экономических исследованиях

Клачкова Ольга Александровна
экономический факультет
МГУ имени М.В. Ломоносова

- GPT vs. GDP: Как большие языковые модели объясняют экономику лучше экономистов (1 попытка)
- Сколько аналитиков нужно, чтобы обработать 10.000 страниц? Один, если у него есть ChatGPT (3 попытки)
- Без труда не вытянешь и рыбку из пруда... если только у тебя нет ChatGPT (5 попыток)



Искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека.

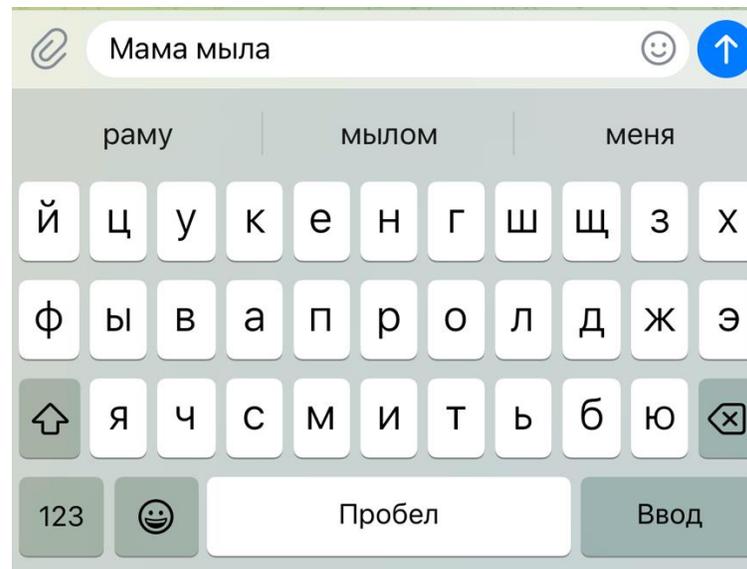
Однотысячерукий экономист

Автоматизированная реализация **технических функций**:

- автоматический сбор больших данных
- подбор литературы (этичный!)
- low-code моделирование / помощь с написанием кода
- оформление: визуализация, работа с качеством текста, ...

Large Language Models ...

- угадывают, какое следующее слово должно идти за уже имеющимся текстом,
- пытаются предсказать следующее слово (игрек) в зависимости от набора подаваемых на вход модели предыдущих слов (иксов)
- и оценивают вероятности возникновения тех или иных слов для продолжения:



Отраслевые исследования: издержки этнической дискриминации для арендодателей

Москва, ЦИАН, 2017 год, 105516 объявлений: характеристики квартиры, географическое расположение, цена аренды и наличие дискриминации в тексте объявления.

1/6 часть объявлений содержит слова 'Asia', 'Caucasus', 'Uzbek', 'Tadjik', 'Kirghiz', 'Armenian', 'Russian', 'Ukrainian', 'Slav' и их производные и являются дискриминирующими.

Найдем квартиры-близнецы – с дискриминацией и без – и сравним цену их аренды \Rightarrow результат – наличие дискриминации снижает цену аренды.

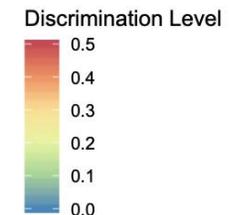
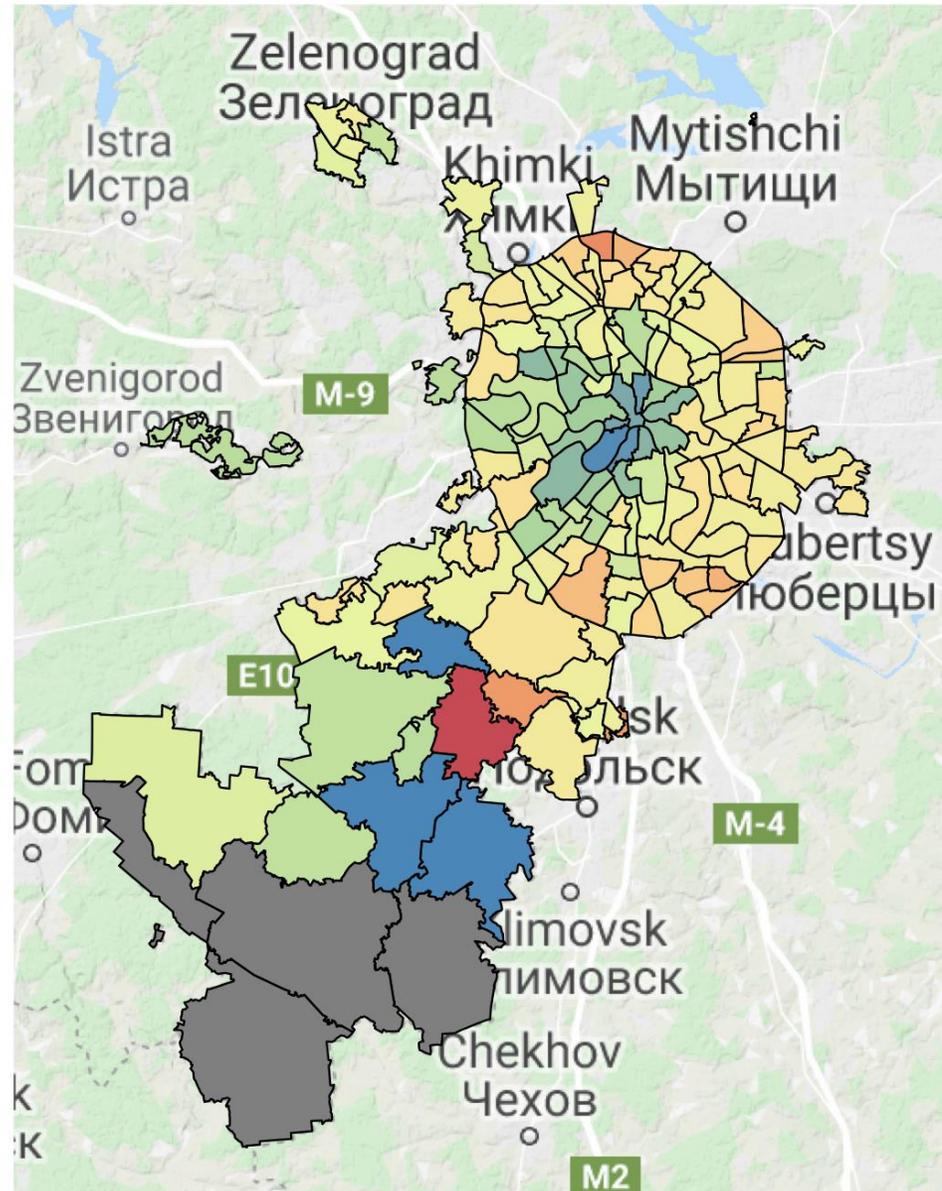
Отраслевые исследования: издержки этнической дискриминации для арендодателей

← 3-комн.кв. • 59,7 м² • 5/5 этаж

АН МакГрупп. Предлагаем в аренду 3-комнатную квартиру в старой центральной части города. После свежего ремонта! Никто не жил! Мебель новая! Дом 1969 г.п., утопает в зелени. Общая площадь 59,7 м2, комнаты смежно-изолированные 15,5+12,5 (смежные), 15,5 м2 изолированная. Окна-новые стеклопакеты, выходят на две стороны, балкон не застеклен. Кухня 7, 2 м2 с газовой панелью. Санузел раздельный, установлены водосчетчики. Спальные места в каждой комнате. Дополнительно оплата по счетчикам. Развитая инфраструктура: детские сады, лицей и гимназия, поликлиники детская и взрослая, магазины, центральный городской рынок, спортивный клуб "Квант", городской стадион, администрация города и налоговая, все в шаговой доступности. Рассмотрим строго славян, желательно семейную пару с детишками, без животных!

Общая информация

Тип жилья	Вторичка
Планировка	Изолированная и совмещенная
Общая площадь	59,7 м²
Жилая площадь	44 м²
Кухня	7,2 м²
Площадь комнат	15,5-15,5+12,5 м²



Отраслевые исследования: спрос на навыки

- Работодателям важно понимать, как соотносить условия, которые они предлагают, и выдвигаемые ими требования.
- Соискателям нужно знать, что лежит в основе формирования их зарплаты.
- Навыки можно рассматривать как инвестиционный товар длительного пользования [Becker, 1964], анализируя выгоды и издержки от их приобретения.

Данные

Автоматизированный сбор данных с сайта HeadHunter: информация о 302 тысячах вакансий в Москве (без опыта):

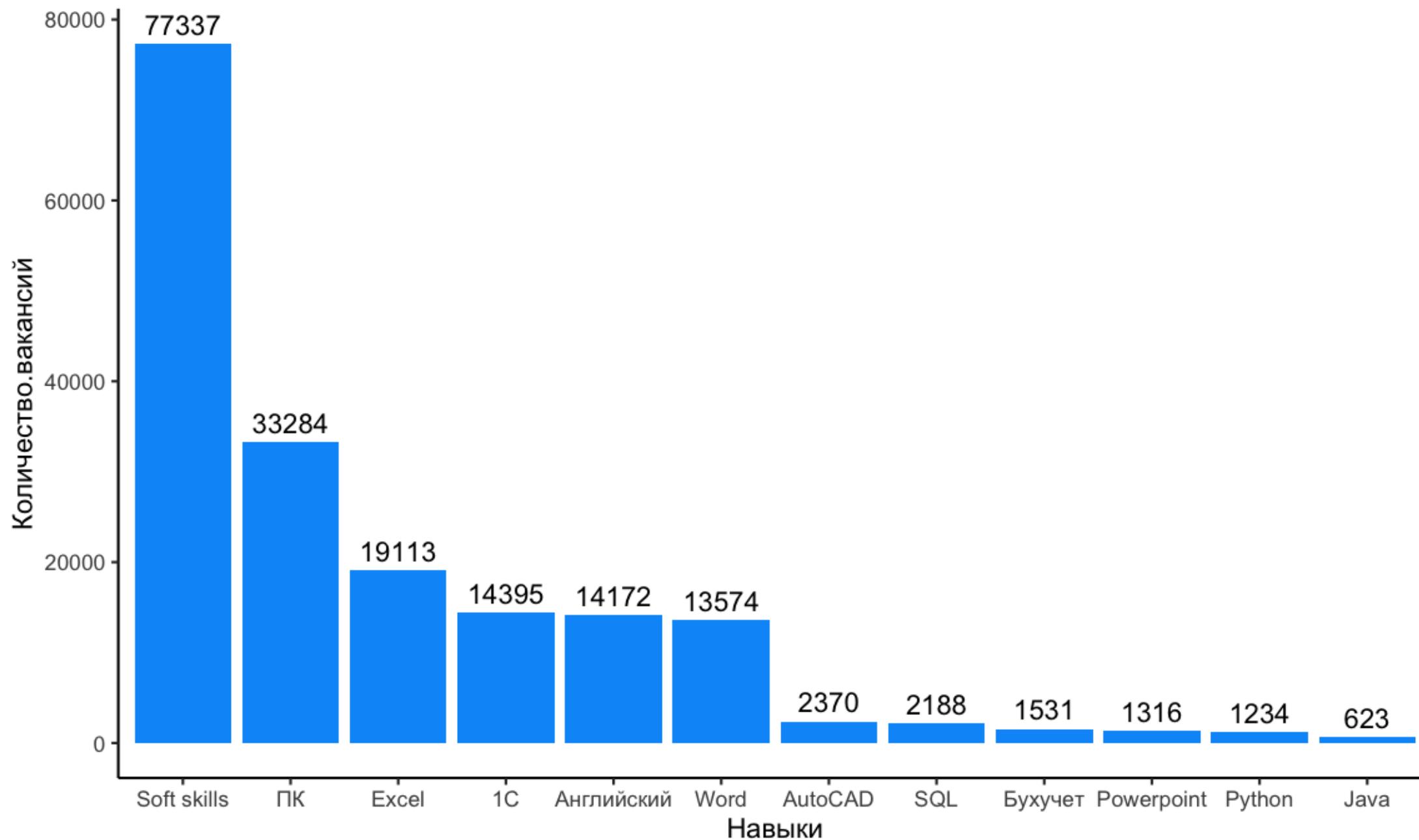
- название вакансии
- дата публикации
- компания
- промежуток предлагаемой заработной платы (от и до)
- полное описание (информация о компании, обязанности сотрудника, требования к кандидату, преимущества и условия работы в компании)
- график работы
- специализация

Название вакансии	Дата публикации	Зарплата (от)	Зарплата (до)	Компания	Адрес	Метро	Описание	График работы	Специализация
Аналитик данных	10.02.2022	50000 Р	80000 Р	ANCOR	Москва, бульвар Энтузиастов, 2	Площадь Ильича	Обязанности: * Автоматизация процессов и отчетов группы; * Требования: * Высшее техническое образование ; * Аналитический склад ума; * Excel – отличное знание Условия: * Официальное оформление по ТК РФ; * Социальный пакет, ДМС , ...	Полный день	Аналитик



Название вакансии	Дата публикации	Зарплата	...	Аналитик	Инженер	...	Высшее образование	Excel	Python	SQL	...	ДМС	...	Полный день
Аналитик данных	10.02.2022	65000 Р	...	1	0	...	1	1	0	0	...	1	...	1

Часто встречающиеся навыки в описании вакансий



Результаты

- Технические навыки положительно влияют на зарплату.
- Влияние soft skills преимущественно косвенное и происходит через усиление влияния технических навыков.
- Навыки, которые реже встречаются в требованиях вакансий, окупаются быстрее.

Макроэкономика: управление инфляционными ожиданиями

Инфляционные ожидания влияют на решения:

- домохозяйств [Carvalho, Nechio, 2014; Dräger, Lamla, 2015; Duca, Kenny, Reuter, 2021]
- фирм [Coibion, Gorodnichenko, Kumar, 2018, Coibion, Gorodnichenko, Ropele, 2020]

Качественная коммуникация позволяет управлять ожиданиями агентов:

- предоставление простой информации об инфляционном таргете и прогнозах по инфляции позволяет снизить инфляционные ожидания агентов [Coibion, Gorodnichenko, Kumar, Pedemonte, 2022; Coibion, Gorodnichenko, Weber, 2022]
- информация, донесенная более простым языком, более эффективна [Carotta, Mello, Ponce, 2023; Mochhoury, 2023]

Коммуникационная политика ЦБ РФ

- работа с профессиональной аудиторией (аналитики, бизнес, научное сообщество) [прозрачность по Dincer, Eichengreen, 2014]
- работа с непрофессиональной аудиторией (население в широком смысле)

Сайт Банка России:

- Доклад о денежно-кредитной политике
- Новости и Пресс-релизы
- Интервью и Выступления

Социальные сети:

- Видеоролики на YouTube-канале («Понятная экономика», «Что почем?»)
- Телеграм-канал
- Группа в VK

Прочие ресурсы:

- Финкульт (ресурс по финансовой грамотности для широкой аудитории)
- Портал «Эконс» (площадка коммуникации внутри профессионального сообщества)

Данные:

- пресс-релизы с января 2015 года
- новости с сентября 2018 года

Пресс-релиз ЦБ: оригинальный текст

Совет директоров Банка России 15 декабря 2023 года принял решение повысить ключевую ставку на 100 б.п., до 16,00% годовых. Текущее инфляционное давление остается высоким. По итогам 2023 года годовая инфляция ожидается вблизи верхней границы прогнозного диапазона 7,0–7,5%. При этом рост ВВП в 2023 году, по оценке Банка России, сложится выше октябрьского прогноза и превысит 3%. Это означает, что отклонение российской экономики вверх от траектории сбалансированного роста во втором полугодии 2023 года оказалось более значительным, чем Банк России оценивал в октябре. В отдельных сегментах кредитного рынка появились признаки замедления активности, однако общие темпы роста кредитования по-прежнему остаются высокими. Повысились инфляционные ожидания населения и ценовые ожидания предприятий. Возвращение инфляции к цели в 2024 году и ее дальнейшая стабилизация вблизи 4% предполагают продолжительный период поддержания жестких денежно-кредитных условий в экономике.

Банк России будет принимать дальнейшие решения по ключевой ставке с учетом фактической и ожидаемой динамики инфляции относительно цели, развития экономики на прогнозном горизонте, а также оценивая риски со стороны внутренних и внешних условий и реакции на них финансовых рынков. Согласно прогнозу Банка России, с учетом проводимой денежно-кредитной политики годовая инфляция снизится до 4,0–4,5% в 2024 году и будет находиться вблизи 4% в дальнейшем.

Динамика инфляции. Текущее инфляционное давление остается высоким. В среднем за октябрь — ноябрь рост цен с поправкой на сезонность составил 10,0% в пересчете на год (по сравнению с 12,2% в III квартале). Годовая инфляция, по оценке на 11 декабря, снизилась до 7,1% после 7,5% в ноябре. Это связано с эффектом высокой базы начала декабря 2022 года, когда произошла внеплановая индексация тарифов на коммунальные услуги. По итогам 2023 года годовая инфляция ожидается вблизи верхней границы прогнозного диапазона 7,0–7,5%.

Устойчивое инфляционное давление в последние месяцы усилилось. Это объясняется более быстрым ростом внутреннего спроса по сравнению с возможностями расширения выпуска товаров и услуг, чем Банк России оценивал ранее. Об этом свидетельствуют данные о динамике экономической активности во втором полугодии 2023 года, включая оперативные данные за IV квартал. В среднем за октябрь — ноябрь показатель базовой инфляции с поправкой на сезонность увеличился до 11,5% в пересчете на год (по сравнению с 9,7% в III квартале). Некоторый вклад в текущий рост цен в последние месяцы также внесли разовые проинфляционные эффекты на отдельных товарных рынках.

Инфляционные ожидания населения и ценовые ожидания предприятий повысились. Инфляционные ожидания профессиональных аналитиков остаются выше цели по инфляции на 2024 год, но на среднесрочную перспективу они закорены вблизи 4%.

Согласно прогнозу Банка России, с учетом проводимой денежно-кредитной политики годовая инфляция снизится до 4,0–4,5% в 2024 году и будет находиться вблизи 4% в дальнейшем.

Денежно-кредитные условия в целом продолжили ужесточаться вслед за повышением ключевой ставки в июле—октябре. Краткосрочные ставки финансового рынка повысились в связи с пересмотром верхучастниками рынка ожидаемой траектории ключевой ставки. На рынке государственного долга средне- и долгосрочные доходности существенно не изменились. Продолжился рост процентных ставок на кредитно-депозитном рынке.

Вслед за ростом депозитных ставок увеличился приток средств населения в кредитные организации, в том числе за счет возврата части наличных денег на банковские счета. Одновременно продолжился переток средств с текущих счетов на срочные депозиты.

В отдельных сегментах кредитного рынка появились признаки замедления активности, однако общие темпы роста кредитования по-прежнему остаются высокими. Замедление необеспеченного потребительского кредитования в последние месяцы происходило в результате как увеличения процентных ставок, так и эффектов ранее принятых макропроденциальных мер. Также замедлилась активность в рыночном сегменте ипотечного кредитования, хотя темпы роста всего ипотечного портфеля были высокими из-за большого объема выданных в рамках государственных льготных программ. Особенно высокой оставалась активность в корпоративном сегменте, где темпы роста кредитования складываются выше октябрьского прогноза Банка России. Повышенный спрос корпоративного сектора на кредиты поддерживается как высокими ценовыми ожиданиями, так и ожиданиями относительно будущей динамики внутреннего спроса.

Экономическая активность. Рост ВВП в III квартале и оперативные данные за IV квартал свидетельствуют о том, что увеличение экономической активности происходит существенно быстрее, чем Банк России ожидал в октябре. Это означает, что отклонение российской экономики вверх от траектории сбалансированного роста во втором полугодии 2023 года оказалось более значительным. Это нашло отражение в показателях текущего роста цен. По оценке Банка России, рост ВВП в 2023 году сложится выше октябрьского прогноза и превысит 3%.

Основной вклад в отклонение российской экономики вверх от траектории сбалансированного роста внес внутренний спрос. Его рост происходит за счет расширения частного спроса при сохранении государственного спроса на высоком уровне. Увеличение потребительской активности поддерживается ростом реальных заработных плат и кредитованием. Значительный рост прибыли компаний и позитивные деловые настроения, в том числе из-за бюджетных стимулов, поддерживают высокий инвестиционный спрос.

Оперативные данные по внешней торговле указывают на то, что импорт несколько снизился по сравнению с пиковыми значениями летних месяцев, в том числе за счет эффектов произошедшего ужесточения денежно-кредитной политики. В то же время снижение экспорта в последние месяцы было более выраженным из-за ухудшения конъюнктуры мировых рынков энергоносителей. В ближайшие кварталы спрос на импорт в рублевом выражении продолжит подстраиваться к уже принятым решениям по ключевой ставке.

На текущем этапе ключевое ограничение на стороне предложения в российской экономике связано с состоянием рынка труда. Согласно данным опросов предприятий, дефицит свободных трудовых ресурсов остается значительным, особенно в обрабатывающих отраслях. Безработица вновь обновила исторический минимум. Невысокая географическая и межотраслевая мобильность рабочей силы является дополнительным структурным ограничением.

Инфляционные риски. На среднесрочном горизонте проинфляционные риски остаются существенными.

В базовом сценарии Банк России исходит из уже принятых решений Правительства по среднесрочной траектории расходов федерального бюджета и бюджетной системы в целом. В случае дополнительного расширения бюджетного дефицита проинфляционные риски вновь возрастут и может потребоваться более жесткая денежно-кредитная политика для возвращения инфляции к цели в 2024 году и ее поддержания вблизи 4% в дальнейшем.

Значимым риском является сохранение инфляционных ожиданий на повышенных уровнях или их дальнейший рост. В этом случае кредитование будет продолжать расширяться ускоренными темпами, а склонность населения к сбережению — снижаться. Как следствие, рост внутреннего спроса будет продолжать опережать возможности расширения предложения и усиливать устойчивое инфляционное давление в экономике. В условиях же ограниченности свободных трудовых ресурсов рост производительности труда может сильнее отставать от роста реальных заработных плат.

Проинфляционные риски связаны с влиянием геополитической напряженности на условия внешней торговли. Усиление внешних торговых и финансовых ограничений может дополнительно ослабить спрос на товары российского экспорта и иметь проинфляционное влияние через динамику валютного курса. Кроме того, усложнение производственных и логистических цепочек или финансовых расчетов в результате внешних ограничений может привести к удорожанию импорта. Значимые краткосрочные проинфляционные эффекты могут также иметь ухудшение перспектив роста мировой экономики и усиление волатильности на мировых товарных и финансовых рынках.

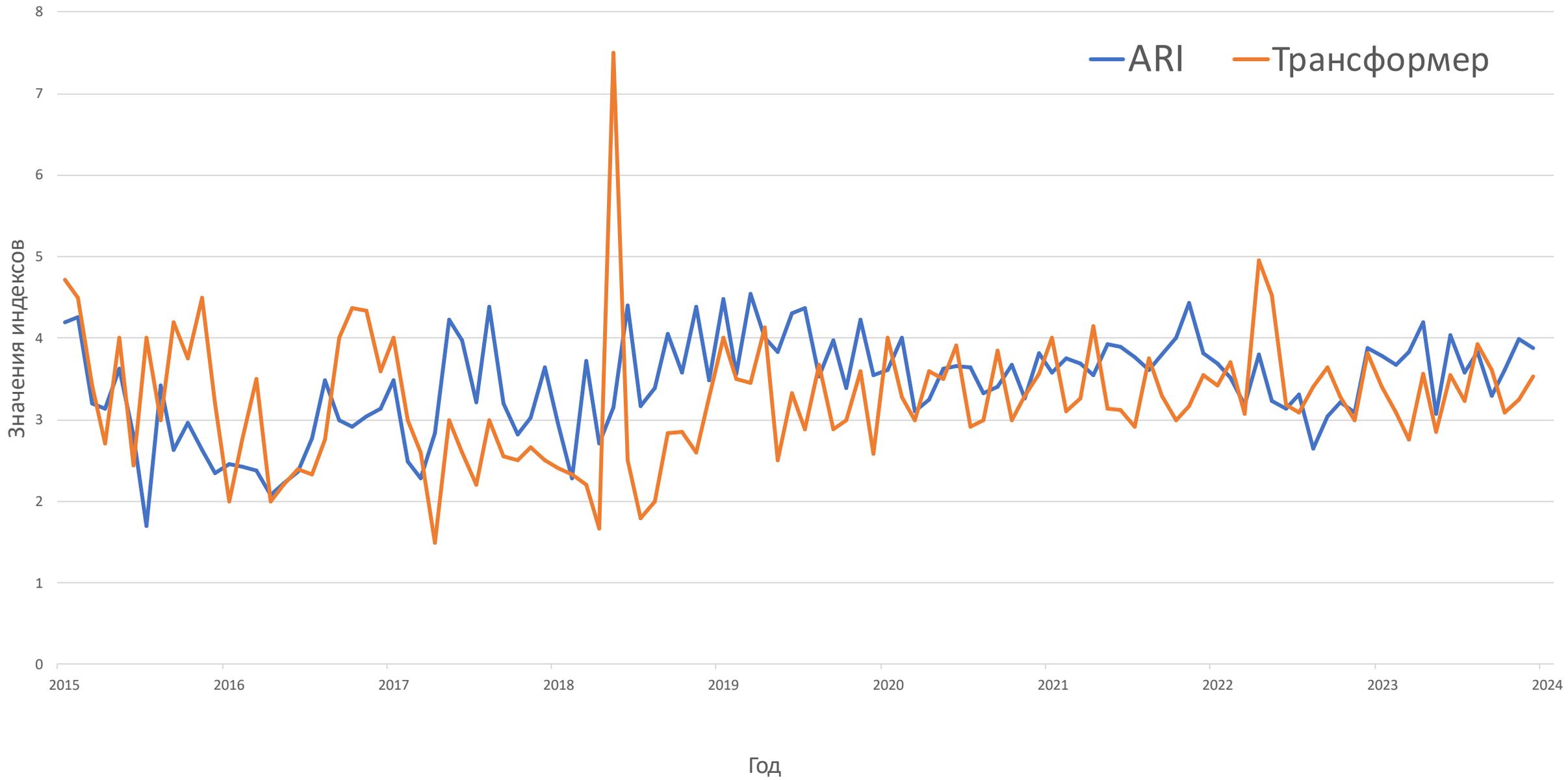
Дезинфляционные риски в первую очередь связаны с более быстрым замедлением внутреннего спроса под влиянием уже реализованного ужесточения денежно-кредитной политики.

Возвращение инфляции к цели в 2024 году и ее дальнейшая стабилизация вблизи 4% предполагают продолжительный период поддержания жестких денежно-кредитных условий в экономике. Банк России будет принимать дальнейшие решения по ключевой ставке с учетом фактической и ожидаемой динамики инфляции относительно цели, развития экономики на прогнозном горизонте, а также оценивая риски со стороны внутренних и внешних условий и реакции на них финансовых рынков.

Сложность коммуникации: удобочитаемость

Удобочитаемость – свойство текстового материала, характеризующее лёгкость восприятия его человеком в процессе чтения.

- Average Readability Index: средняя длина слов и средняя длина предложений [Flesch, 1948; Kincaid et al., 1975; Оборнева, 2006]
- Нейронная сеть по 44 лингвистическим характеристикам [Evstigneeva, Sidorovski, 2021]:
 - 12 синтаксических («структура предложений»);
 - 16 лексических («характеристики отдельных слов»);
 - 7 морфологических («веса отдельных частей речи или структуры слова»);
 - 1 фонетическая («благозвучие текста»);
 - 5 семантических («значения составляющих текст слов»);
 - 3 дискурсивные («внутренние связи текста и параметры функциональных стилей речи»)



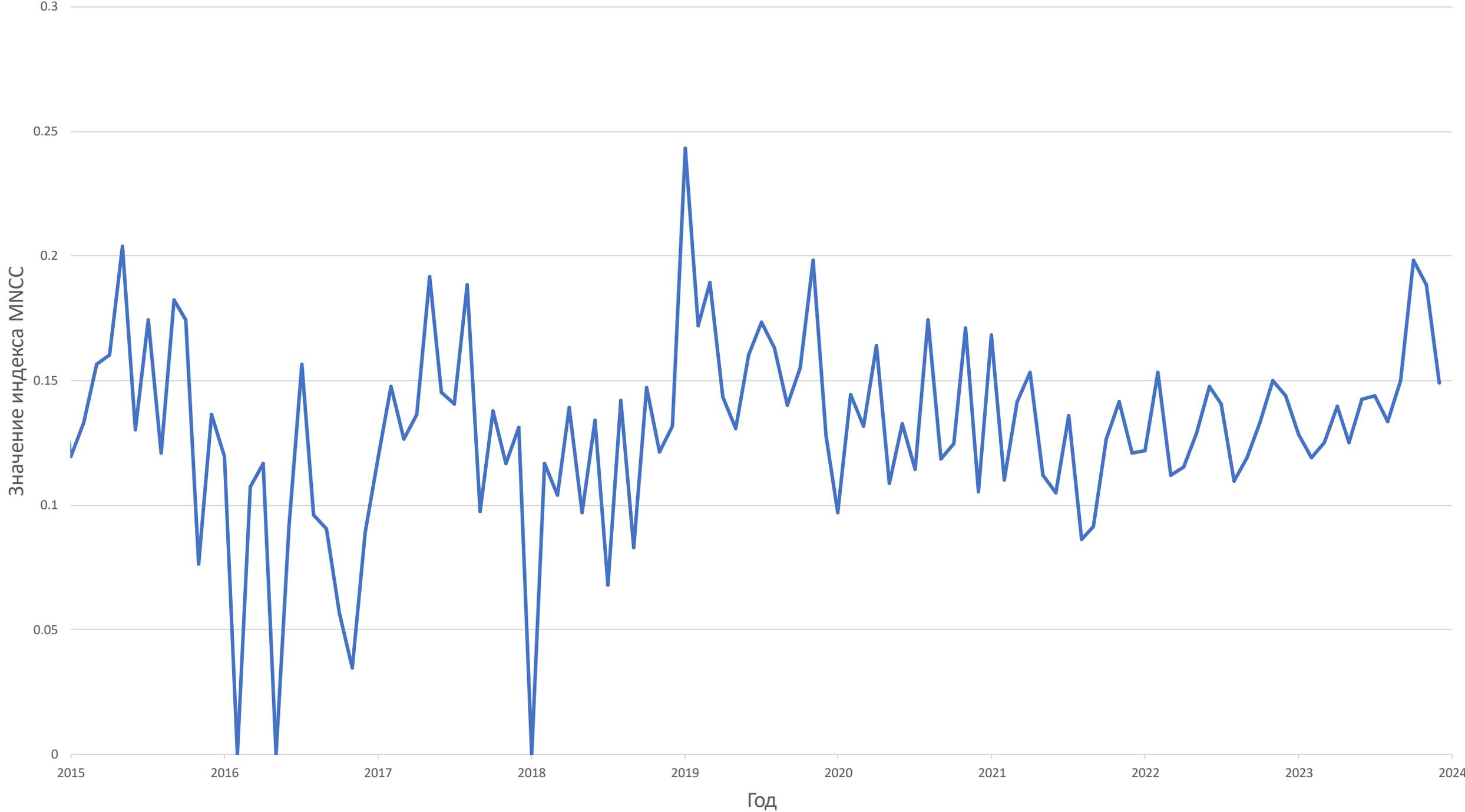
Сложность коммуникации: концептуальная

Концептуальная сложность – отражает насыщенность текста терминами.

McMahon-Naylor Conceptual Complexity (MNCC) [McMahon, Naylor, 2023] – индекс концептуальной сложности. Он учитывает:

- долю экономических терминов (жаргонов) в общем числе слов;
- разнообразие жаргонов в рамках определенной экономической темы;
- число охватываемых в тексте тем.

Термины были переведены на русский язык, всего 10 тем и около 350 терминов.



Тональность коммуникации

- «голубиная» тональность: идеи смягчения денежно-кредитной политики
- «ястребиная» тональность: идеи ужесточения монетарной политики

Как оценить?

- экспертная оценка [Ehrmann, Fratzscher, 2007; Hubert, 2016]
- по соседним к ключевым словам в предложениях [Carotta, Mello, Ponce, 2023]
- генеративным ИИ, GPT-4 [Carotta, Mello, Ponce, 2023; Hansen, Kazinnik, 2023; Smales, 2023; Woodhouse, Charlesworth, 2023]

A You

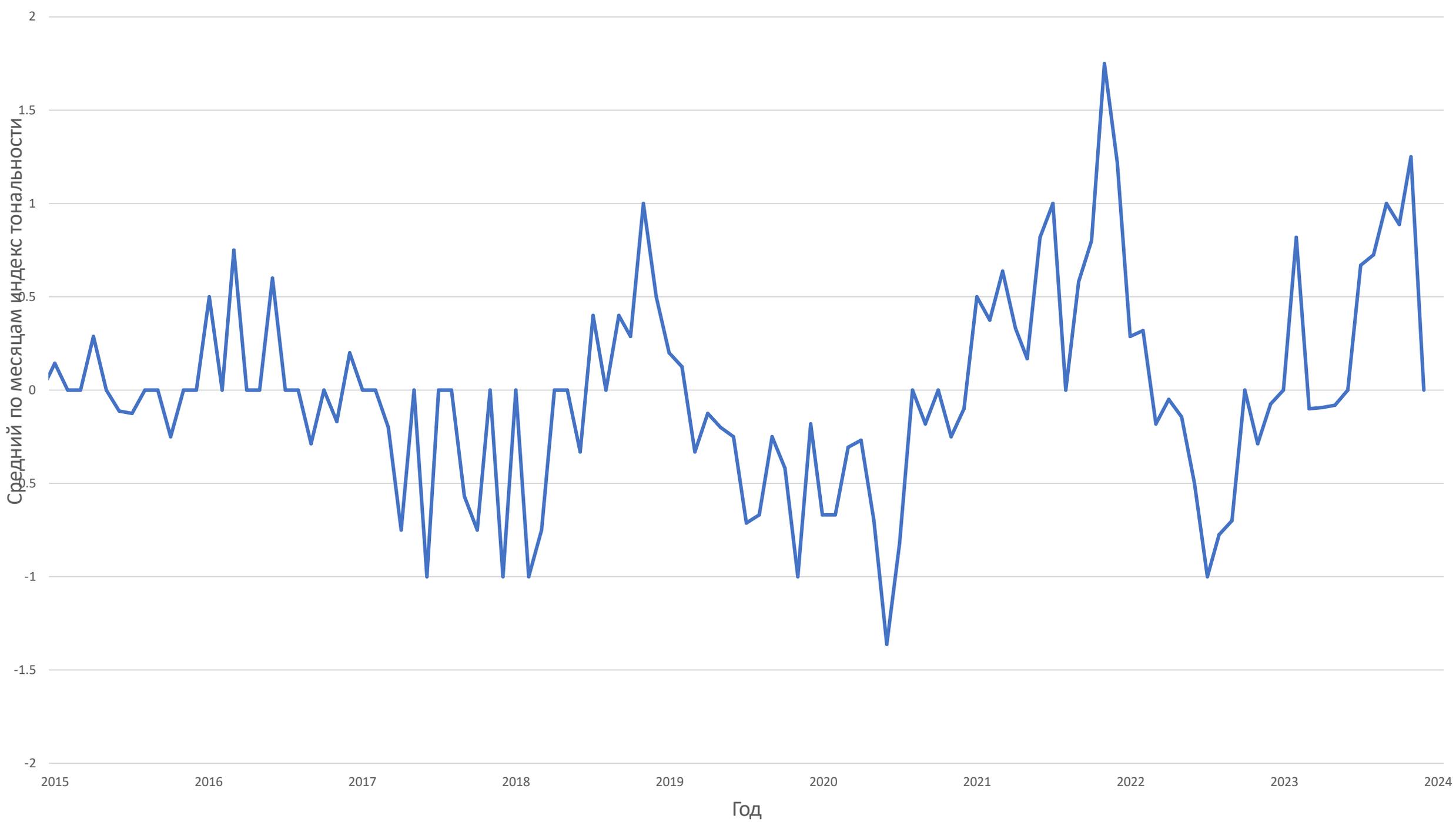
Представьте, что вы - научный сотрудник, работающий в Банке России. Ваша задача - отнести следующую статью к одной из пяти категорий с точки зрения монетарной политики (голубиная, в основном голубиная, нейтральная, в основном ястребиная, ястребиная). Если в статье недостаточно информации для классификации, поставьте «-». Свой ответ поясни.

В феврале индикаторы денежно-кредитных условий (ДКУ) по-прежнему ужесточались, реагируя на предпринятые во второй половине 2023 года повышения ключевой ставки и на сигнал Банка России о поддержании в течение продолжительного времени жестких ДКУ. Доходности облигаций федерального займа в феврале выросли. Показатели инфляционных ожиданий снизились, но остались повышенными. Кредитная активность стала более сдержанной, чем в III–IV кварталах прошлого года. Годовые темпы роста денежных агрегатов несколько замедлились, но сохранились выше значений 2017–2019 годов. Более подробно читайте в информационно-аналитическом комментарии «Денежно-кредитные условия и трансмиссионный механизм денежно-кредитной политики».

ChatGPT

Ответ: ястребиная.

Пояснение: В статье описывается ситуация, когда Банк России поддерживает ужесточенные денежно-кредитные условия (ДКУ), включая повышения ключевой ставки и сигналы о жесткой монетарной политике на продолжительный период. Увеличение доходности облигаций и сдерживание кредитной активности также указывают на приоритет борьбы с инфляцией и контроль над экспансивной финансовой активностью. Эти меры характерны для ястребиной монетарной политики, главной целью которой является сдерживание инфляции, даже за счет потенциального замедления экономического роста.



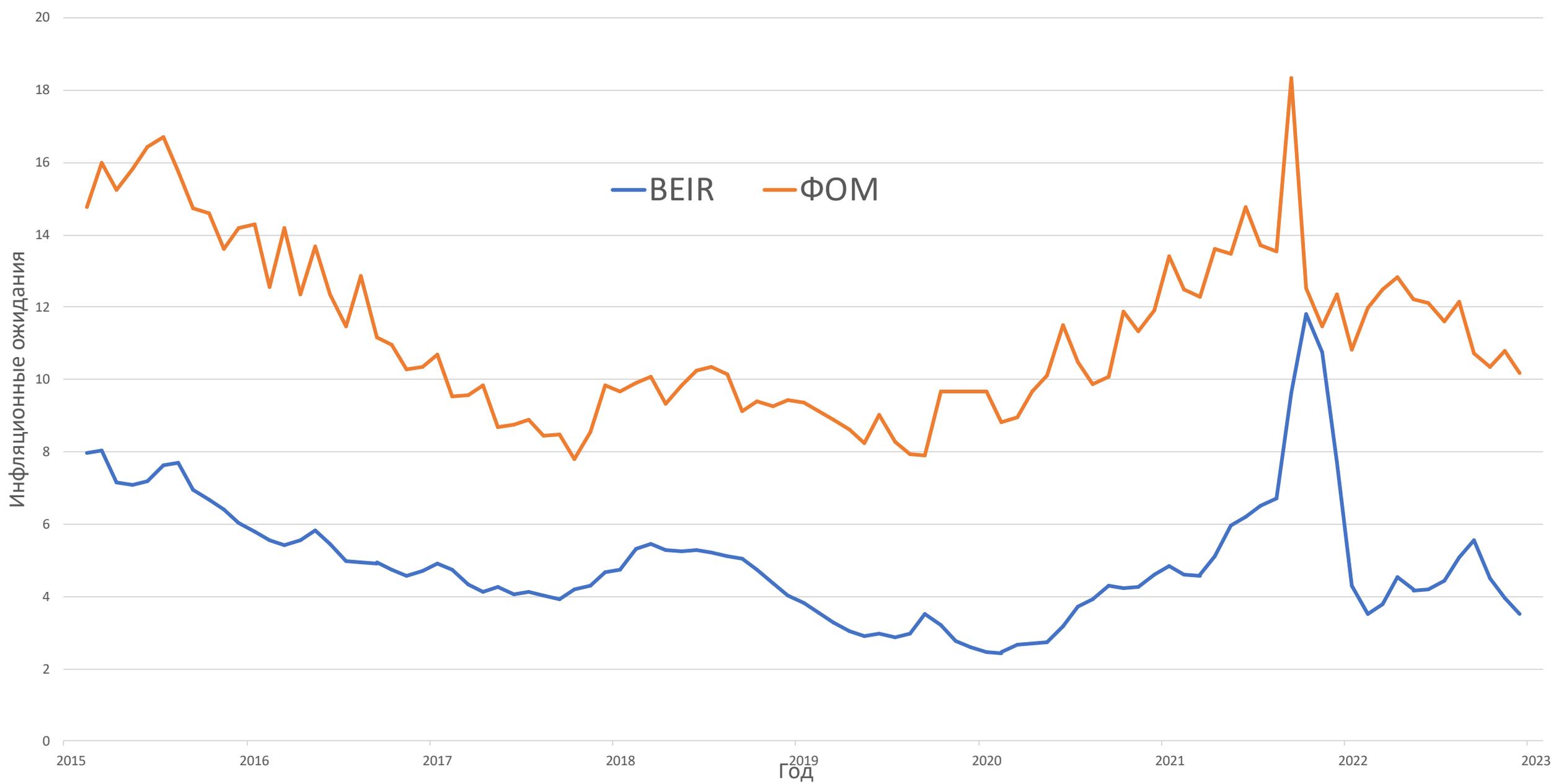
Данные

ИО населения

- опросы ФОМ (месячные данные);
- ответы на вопрос: «Как, по вашему мнению, в целом будут меняться цены в следующие 12 месяцев (год)?»

ИО рынка [Жемков, Кузнецова, 2017]

- вмененная инфляция (BEIR), рассчитанная через ОФЗ (дневные данные);
- для расчета BEIR берем разность доходностей двух облигаций: индексируемой на инфляцию [ОФЗ-520001-ИН] и номинальной [ОФЗ-26215-ПД]



Регрессия: выводы

- Тональность оказывает **слабое положительное влияние** на динамику инфляционных ожиданий рынка (BEIR), удобочитаемость, измеренная двумя различными способами, это влияние не усиливает (**профессиональная аудитория**).
- Тональность **не оказывает значимого влияния** на инфляционные ожидания населения (**непрофессиональная аудитория**) и волатильность инфляционных ожиданий рынка.
- Потенциальное объяснение: информированные агенты, говорящие с центральным банком «на одном языке», могут понимать, что свобода маневра ЦБ РФ при применении традиционных мер денежно-кредитной политики сокращена в условиях санкционного давления и ограничений на мобильность капитала.

Экспериментальная оценка влияния удобочитаемости на ожидания [Картаев, Клачкова, 2025]

- Лабораторный эксперимент
- 274 респондента
- Контрольная группа получала в качестве коммуникации оригинальные пресс-релизы Банка России
- Экспериментальная группа получала эти же пресс-релизы, преобразованные с целью улучшения удобочитаемости
- Улучшение осуществлялось при помощи большой языковой модели
- Два отдельных эксперимента: пресс-релиз, посвященный снижению ставки; пресс-релиз, посвященный увеличению ставки

Пресс-релиз ЦБ: упрощенный текст

Совет директоров Банка России решил повысить ключевую ставку до 16%. Это означает, что кредиты станут дороже, чтобы сдержать инфляцию, которая в 2023 году ожидается на уровне 7-7,5%.

Инфляция, или рост цен на товары и услуги, остается высокой. В октябре и ноябре цены выросли на 10% по сравнению с прошлым годом. В декабре инфляция снизилась до 7,1%, благодаря тому, что в прошлом году в это время резко подняли тарифы на коммунальные услуги.

Экономический рост в России оказался выше прогнозов: ВВП в 2023 году вырастет более чем на 3%, благодаря высокому спросу на товары и услуги, росту зарплат и кредитования. Однако на рынке труда остро не хватает рабочих рук.

Для стабилизации инфляции на уровне около 4% в 2024 году Банк России будет и дальше держать высокие процентные ставки. Это замедлит рост кредитования и, следовательно, расходы людей и компаний.

Ожидается, что ситуация с высокими ценами продлится, так как спрос в России растет быстрее, чем производство. Это означает, что Банк России продолжит следить за инфляцией и экономикой, чтобы принимать решения по ключевой ставке и не дать инфляции сильно вырасти.

Экспериментальная оценка влияния удобочитаемости на ожидания [Картаев, Клачкова, 2025]

Результаты:

- При повышении ставки упрощенная коммуникация снижает инфляционные ожидания на 1,8 п.п.
- При понижении ставки тип коммуникации не важен для ожиданий

- Препод понял, что данные сгенерированы, когда увидел, что уровень инфляции в США равен 200%, а курс доллара — 3 картошки
- Штирлиц был близок к провалу, но студент был ещё ближе, когда его спросили, почему все числа в выборке делятся на 5 без остатка.

Расходы на R&D (млрд руб.)

