



Национальный исследовательский  
университет «Высшая школа экономики»

Факультет  
экономических наук

Москва, 2025

# Оценка прогностического потенциала больших языковых моделей для наукастинга ключевых макроэкономических показателей России

Выполнила:

**Лашина Мария Дмитриевна**

Научный руководитель:

**Станкевич Иван Павлович**, к.э.н., доцент Департамента прикладной экономики

Рецензент:

**Шибанов Олег Константинович**, к.ф.-м.н., PhD, профессор РЭШ

## Введение: Предпосылки и актуальность



- Прогнозы важны для государственного управления и планирования всех экономических субъектов.
- Сложность прогнозирования экономических показателей обусловлена редкостью и разнообразием структурных разрывов, зависящих от множества экзогенных факторов.
- Применение продвинутых моделей целесообразно для повышения точности прогнозов.
- В ML происходит смена парадигмы – переход к фундаментальным моделям. Это большие нейронные сети, предварительно обученные на больших объемах разнообразных данных, которые демонстрируют способность к обобщению знаний и часто превосходят модели, обученные на конкретные задачи.

## Введение: Мотивация и постановка задачи

### Наблюдение:

LLM успешно предсказывают временные ряды, которых не было в обучении (Bommasani, 2022).

### Цель:

Оценка прогностической эффективности LLM в задаче прогнозирования динамики экономических показателей Российской Федерации.

SARIMA

**VS.**

LLM:

- Chronos
- Lag-Llama

### Гипотезы:

- 1 RMSE прогнозов LLM ниже, чем у SARIMA
- 2 Ошибка прогнозов LLM ниже для данных более высокой частоты
- 3 RMSE прогнозов LLM ниже для категории финансовых показателей

## Обзор литературы: Архитектура и обучение LLM

1. Предобучение: LLM учится предсказывать следующее слово на текстах Интернета, предварительно отфильтрованных.
2. Дообучение: LLM обучается взаимодействию в качестве ассистента на написанных вручную парах текстов, содержащих инструкцию и эталонный ответ.
3. (Опционально) Обучение с подкреплением: Обучение LLM деталям для улучшения качества ответов на реальных запросах пользователей.



## Обзор литературы: Архитектура и обучение LLM

1. Предобучение: LLM учится предсказывать следующее слово на текстах Интернета, предварительно отфильтрованных.
2. Дообучение: LLM обучается взаимодействию в качестве ассистента на написанных вручную парах текстов, содержащих инструкцию и эталонный ответ.
3. (Опционально) Обучение с подкреплением: Обучение LLM деталям для улучшения качества ответов на реальных запросах пользователей.



Несмотря на то, что основное применение LLM – работа с текстами, эти модели оперируют числами. Также LLM обрабатывают последовательные структуры. Поэтому применение LLM для обработки временных рядов является естественным и перспективным направлением исследований.

## Обзор литературы: LLM и временные ряды

Этап/ подход	Описание	Пример модели
Входной текст/ Промптирование	Числовой временной ряд передается в LLM как обычный текст	PromptCast (Xue & Salim, 2023)
	Специальная обработка чисел с учетом токенизации LLM	LLMTime (Gruver, Finzi, Qiu, 2023)
Токенизация/ Квантизация	Обучение VQ-VAE кодера для получения скрытых представлений входных данных	Auto-TTE (Chung, et al., 2023)
	Кластеризация K-Means для индексной токенизации	SpeechGPT (Wen, и др., 2022)
	Частотный спектр как общий словарь	FreqTST (Anonymous, 2023)
	Масштабирование и квантование	Chronos (Ansari, et al., 2024)
	Предопределенные текстовые категории	TDML (Yu, et al., 2023)
Энкодер/ Выравнивание	Минимизация контрастных или других потерь	ETP (Liu, Wan, Cheng, 2023)
	Интеграция LLM в качестве вспомогательных слоев	Lag-Llama (Rasul, et al., 2024)
Декодер/ VLM	Обучение на парных данных с изображениями	ImageBind (Girdhar, et al., 2023)
	Представление временных рядов как графиков	Insight Miner (Zhang, et al., 2023)

## Обзор литературы: LLM и временные ряды - Промптирование

### Пример. LLMTime

#### Временной ряд

0.123, 1.23, 12.3, 123.0 ...



#### Масштабирование

$\tilde{x} = x^{-b}/a$ , где  $b = \min(x) - \beta(\max(x) - \min(x))$ ,  $a$  – это  $\alpha$ -процентиль масштабированных на  $b$  значений

Пусть  $\alpha = 0.9$ ,  $\beta = 0.3$ , тогда:

0.2308, 0.2377, 0.3069, 1



#### Токенизация

- добавляем пробелы между каждой цифрой
- наблюдения отделяем запятой
- количество цифр после запятой фиксировано

"0 2 3 , 0 2 4 , 0 3 1 , 1 0 0"

# Обзор литературы: LLM и временные ряды

Этап/ подход	Описание	Пример модели
Входной текст/ Промптирование	Числовой временной ряд передается в LLM как обычный текст	PromptCast (Xue & Salim, 2023)
	Специальная обработка чисел с учетом токенизации LLM	LLMTime (Gruver, Finzi, Qiu, 2023)
Токенизация/ Квантизация	Обучение VQ-VAE кодера для получения скрытых представлений входных данных	Auto-TTE (Chung, et al., 2023)
	Кластеризация K-Means для индексной токенизации	SpeechGPT (Wen, и др., 2022)
	Частотный спектр как общий словарь	FreqTST (Anonymous, 2023)
	Масштабирование и квантование	Chronos (Ansari, et al., 2024)
	Предопределенные текстовые категории	TDML (Yu, et al., 2023)
Энкодер/ Выравнивание	Минимизация контрастных или других потерь	ETP (Liu, Wan, Cheng, 2023)
	Интеграция LLM в качестве вспомогательных слоев	Lag-Llama (Rasul, et al., 2024)
Декодер/ VLM	Обучение на парных данных с изображениями	ImageBind (Girdhar, et al., 2023)
	Представление временных рядов как графиков	Insight Miner (Zhang, et al., 2023)



## Обзор литературы: LLM и временные ряды - Квантизация

### Пример. Chronos

#### Временной ряд

0.123, 1.23, 12.3, 123.0 ...



#### Масштабирование

По среднему:  $\tilde{x} = x/m$ , где  $m$  – среднее исторического временного ряда

0.0036, 0.0360, 0.3600, 3.6003



#### Квантизация

Выбирается стратегия бинования, например, количество бинов (размер словаря токенов) 4096, максимальная граница масштабированных данных 15, минимальная -15, тогда ширина

бина  $\frac{15 - (-15)}{4096} = 0.0073$ .

[0, 0.0073]=0.0037, [0.0293, 0.0366]=0.0329, [0.3589, 0.3662]=0.3625, [3.6548, 3.6621]=3.6584

## Обзор литературы: LLM и временные ряды

Этап/ подход	Описание	Пример модели
Входной текст/ Промптирование	Числовой временной ряд передается в LLM как обычный текст	PromptCast (Xue & Salim, 2023)
	Специальная обработка чисел с учетом токенизации LLM	LLMTime (Gruver, Finzi, Qiu, 2023)
Токенизация/ Квантизация	Обучение VQ-VAE кодера для получения скрытых представлений входных данных	Auto-TTE (Chung, et al., 2023)
	Кластеризация K-Means для индексной токенизации	SpeechGPT (Wen, и др., 2022)
	Частотный спектр как общий словарь	FreqTST (Anonymous, 2023)
	Масштабирование и квантование	Chronos (Ansari, et al., 2024)
	Предопределенные текстовые категории	TDML (Yu, et al., 2023)
Энкодер/ Выравнивание	Минимизация контрастных или других потерь	ETP (Liu, Wan, Cheng, 2023)
	Интеграция LLM в качестве вспомогательных слоев	Lag-Llama (Rasul, et al., 2024)
Декодер/ VLM	Обучение на парных данных с изображениями	ImageBind (Girdhar, et al., 2023)
	Представление временных рядов как графиков	Insight Miner (Zhang, et al., 2023)

## Обзор литературы: LLM и временные ряды - Выравнивание

### Пример. Lag-Llama

#### Временной ряд

0.123, 1.23, 12.3, 123.0 ...



#### Масштабирование

Приведение к стандартному нормальному распределению, параметры определяются по окну  
-0.71, 0.71, 14.85, 156.27



#### Эмбеддинги – векторы, содержащие смысл токенов

- L лагов
- показатели времени-даты
- среднее и отклонение окна

Пусть L равно 2, из показателей времени-даты будет год.  
(-), (-), (-0.71, 0.71, 2017, 0.68, 0.78), (0.71, 14.85, 2018, 0.68, 0.78)

## Обзор литературы: LLM и временные ряды

Этап/ подход	Описание	Пример модели
Входной текст/ Промптирование	Числовой временной ряд передается в LLM как обычный текст	PromptCast (Xue & Salim, 2023)
	Специальная обработка чисел с учетом токенизации LLM	LLMTime (Gruver, Finzi, Qiu, 2023)
Токенизация/ Квантизация	Обучение VQ-VAE кодера для получения скрытых представлений входных данных	Auto-TTE (Chung, et al., 2023)
	Кластеризация K-Means для индексной токенизации	SpeechGPT (Wen, и др., 2022)
	Частотный спектр как общий словарь	FreqTST (Anonymous, 2023)
	Масштабирование и квантование	Chronos (Ansari, et al., 2024)
	Предопределенные текстовые категории	TDML (Yu, et al., 2023)
Энкодер/ Выравнивание	Минимизация контрастных или других потерь	ETP (Liu, Wan, Cheng, 2023)
	Интеграция LLM в качестве вспомогательных слоев	Lag-Llama (Rasul, et al., 2024)
Декодер/ VLM	Обучение на парных данных с изображениями	ImageBind (Girdhar, et al., 2023)
	Представление временных рядов как графиков	Insight Miner (Zhang, et al., 2023)

# Обзор литературы: LLM и временные ряды - VLM

## Пример. Insight-Miner

### Временной ряд

### Выделение тренда, сглаживание

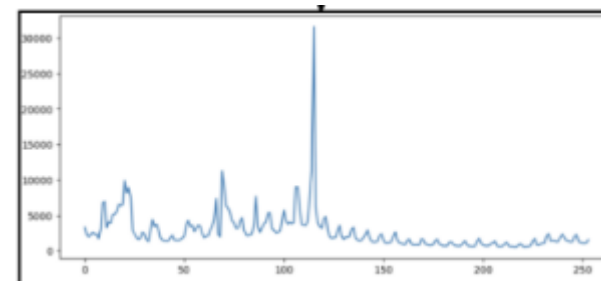
Если сезонность присутствует, то тренд выделяется с помощью Seasonal-Trend Decomposition  
Если сезонность отсутствует, то применяется сглаживание гауссиановским процессом

### Формирование промпта

Преобразованный временной ряд округляется до 1 знака после запятой и передается в GPT-4, который генерирует описание тренда

### Входные данные для VLM

- изображение исходного временного ряда
- описание тренда, сгенерированного GPT-4



**Description:** The trend initially increases, then decreases significantly, stabilizes briefly, and finally shows a slight increase.

# Обзор литературы: LLM в прогнозировании временных

Этап/ подход	Описание	Пример модели
Входной текст/ Промптирование	Числовой временной ряд передается в LLM как обычный текст	PromptCast (Xue & Salim, 2023)
	Специальная обработка чисел с учетом токенизации LLM	LLMTime (Gruver, Finzi, Qiu, 2023)
Токенизация/ Квантизация	Обучение VQ-VAE кодера для получения скрытых представлений входных данных	Auto-TTE (Chung, et al., 2023)
	Кластеризация K-Means для индексной токенизации	SpeechGPT (Wen, и др., 2022)
	Частотный спектр как общий словарь	FreqTST (Anonymous, 2023)
	Масштабирование и квантование	Chronos (Ansari, et al., 2024)
	Предопределенные текстовые категории	TDML (Yu, et al., 2023)
Энкодер/ Выравнивание	Минимизация контрастных или других потерь	ETP (Liu, Wan, Cheng, 2023)
	Интеграция LLM в качестве вспомогательных слоев	Lag-Llama (Rasul, et al., 2024)
Декодер/ VLM	Обучение на парных данных с изображениями	ImageBind (Girdhar, et al., 2023)
	Представление временных рядов как графиков	Insight Miner (Zhang, et al., 2023)

## Обзор литературы: Вклад этого исследования

- 1 LLM для прогнозирования макроэкономики России
- 2 Эффективность LLM для показателей разной частоты
- 3 Качество прогнозов LLM для показателей разных категорий: реальный сектор, финансовые, опережающие

# Модели

## Статистическая модель

**SARIMA** – модель авторегрессии  $AR(p)$  и скользящего среднего  $MA(q)$  с учетом нестационарности и сезонности.

Модель SARIMA была построена с помощью автоматического подбора порядка интегрирования по критерию KSS, порядка сезонного интегрирования с помощью OCSB теста, количества лагов по информационным критериям AIC.

## LLM

### Chronos

- Архитектура: T5 из-за способности к масштабированию и наличия разнообразных конфигураций от 16 миллионов до 11 миллиардов параметров
- Работа с временными данными: сначала нормализация, затем преобразование в фиксированные интервалы
- Данные в обучении: данные различных секторов экономики и науки, включая энергетическую отрасль, транспортную инфраструктуру, сферу здравоохранения, розничную торговлю, интернет-услуги, метеорологические наблюдения и финансовые рынки. Данные разнородны по частоте: от пятиминутных интервалов до годовых периодов

### Lag-Llama

- Архитектура: Llama 2, включающая приблизительно 2,5 миллиона параметров
- Работа с временными рядами: сначала нормализация, применяемая на уровне окна данных, затем токенизация, включающая извлечение признаков запаздывания, основанных на исторических значениях временного ряда с различными временными задержками
- Данные: данные шести различных областей: энергетика, транспорт, экономика, природа, качество воздуха и облачные вычисления. Данные различны по длине прогноза, длине ряда и частоте.



## Данные: сбор и обработка

Категория переменных	Примеры переменных			Количество переменных	Источники данных
	Название	Частота	Период		
Опережающие	Диффузный индекс цен на выпускаемую продукцию, промышленность, фактические изменения (доля предприятий с растущим за 1 месяц показателем)	Y, Q, M	1992-2024	16	Журнал «Российский экономический барометр»
	Диффузный индекс портфеля заказов, промышленность, фактические изменения (доля предприятий с растущим за 1 месяц показателем)	Y, Q, M	1992-2024		
	Диффузный индекс цен на выпускаемую продукцию, промышленность, ожидаемые изменения (доля предприятий с растущим за 3 месяца показателем)	Y, Q, M	1992-2024		
Реальный сектор	Реальный ВВП	Y, Q	2011-2024	12	• ЕАЭСД НИУ ВШЭ • Росстат
	Индекс реальных денежных доходов	Y, Q, M	1992-2024		
	Уровень безработицы	Y, Q, M	1994-2023		
Финансы	ИПЦ	Y, Q, M	1992-2024	6	• ЕАЭСД НИУ ВШЭ • МОЕХ
	Курс доллара	Y, Q, M	1993-2023		
	Индекс РТС	Y, Q, M	1995-2025		

- Разбиение данных на обучающую и тестовую часть в соотношении 9:1
- Данные стандартизировались (среднее и отклонение определялись только на обучающей выборке)

## Результаты: Действительно ли LLM лучше прогнозируют временные ряды? (1/3)

Значения ошибок RMSE для LLM приведены по отношению к RMSE прогнозов SARIMA, для учета специфики каждого ряда.

Модель	RMSE на тестовой выборке			
	Медиана	Стандартное отклонение	Минимальное значение	Максимальное значение
Chronos	1.04	0.58	0.07	3.87
Lag-Llama	1.08	1.04	0.33	6.98



Гипотеза 1 о более низкой RMSE LLM по сравнению с SARIMA **не подтвердилась**.

Результаты несколько отличаются от заявленных в других исследованиях:

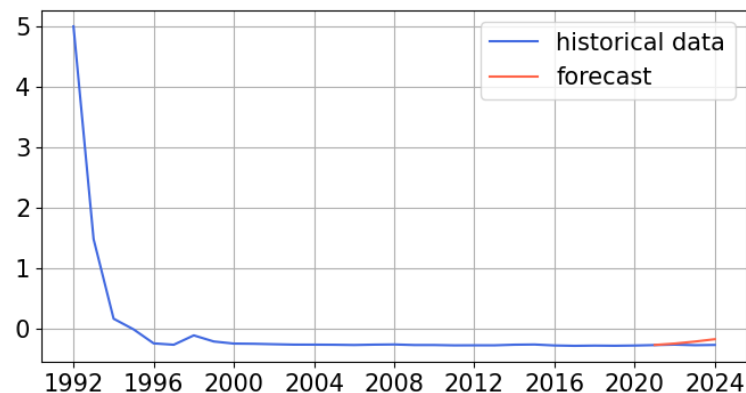
- Сравнение качества прогнозов MIDAS и Lag-LIAMA для ежедневной инфляции стран Еврозоны (Bahelka & De Weerd, 2024) было показано незначительное превосходство LLM.
- Оценка качества нескольких LLM против BVAR и DFM на 120 месячных макропоказателях США (Carriero, Pettenuzzo, & Shekhar, 2025) показала сопоставимую точность прогнозов, однако превосходство над бенчмарком AR(1) не оказалось статистически значимым. Также было отмечена более высокая стабильность прогнозов статистических моделей по сравнению с LLM.

Результаты: Действительно ли LLM лучше прогнозируют временные ряды? (2/3)

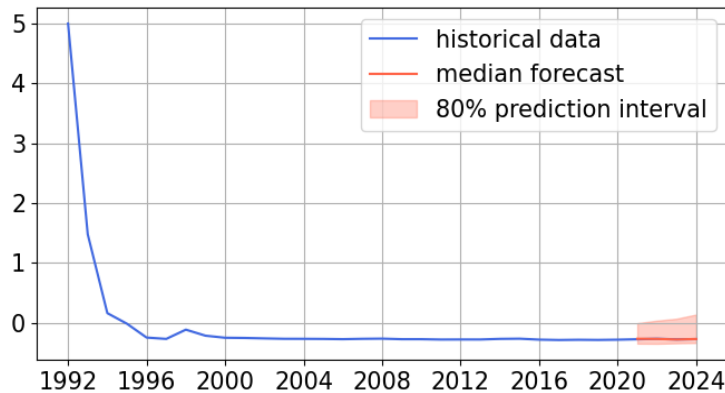
	Показатель	Частота	Категория	RMSE LLM/ RMSE SARIMA
Показатели с лучшими прогнозами Chronos	ИПЦ	Y	Финансовый	0,07
	Индекс реальных денежных доходов	Q	Реал. сектор	0,63
	Диффуз. инд. з/п, промышленность, факт.	Q	Опережающий	0,63
Показатели с лучшими прогнозами Lag-Llama	Цена MOEX	M	Финансовый	0,33
	ИПЦ	Y	Финансовый	0,49
	Индекс дефлятор ВВП	Q	Реал. сектор	0,50
Показатели с худшими прогнозами Chronos	Индекс промыш. производства по ОКВЭД2 с сезон. корр.	M	Реал. сектор	3,87
	Индекс реальных ден. доходов с сезон. корр.	M	Реал. сектор	3,63
	Уровень безработицы	Q	Реал. сектор	2,69
Показатели с худшими прогнозами Lag-Llama	Индекс реальных ден. доходов с сезон. корр.	M	Реал. сектор	6,98
	Индекс реальных инвестиций в основной капитал	M	Реал. сектор	4,19
	Индекс реальных инвест. в основ. капитал с сезон. корр.	M	Реал. сектор	3,96

## Результаты: Действительно ли LLM лучше прогнозируют временные ряды? (<sup>3</sup>/<sub>3</sub>)

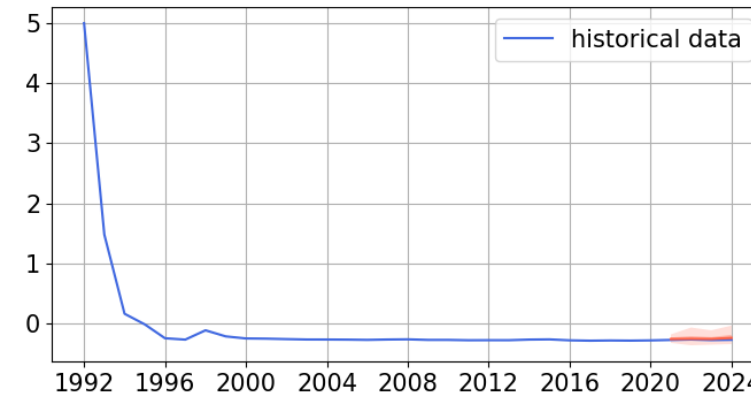
ИПЦ, годовая частота, финансовый показатель



Прогноз SARIMA

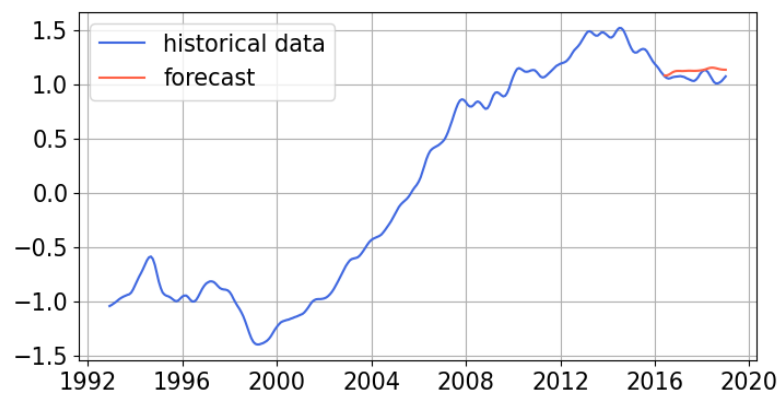


Прогноз Chronos

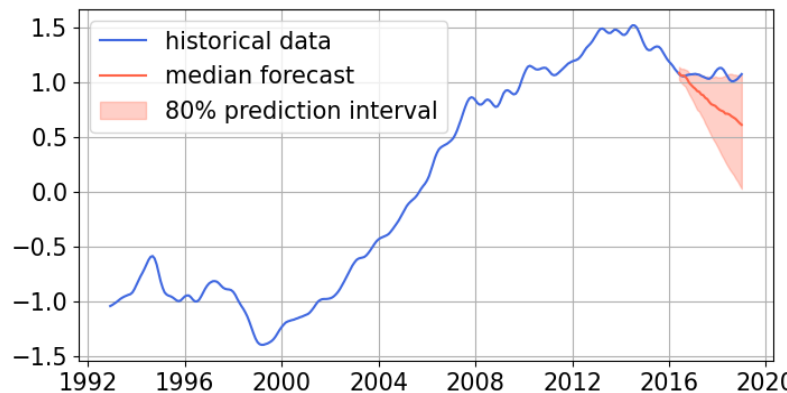


Прогноз Lag-Llama

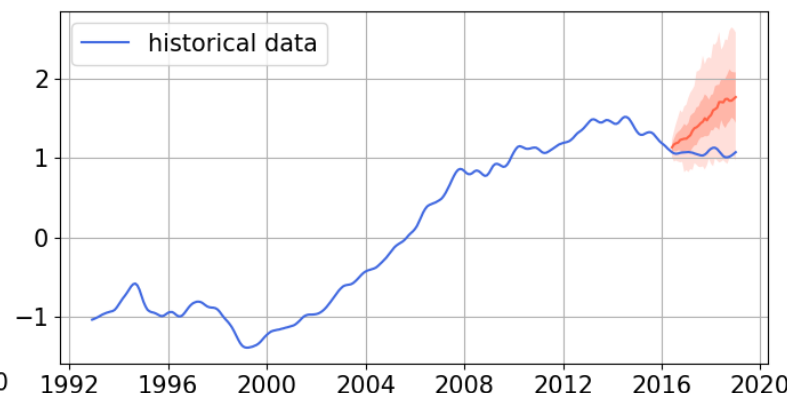
Индекс реальных ден. доходов с сезон. корр., месячная частота, показатель реального сектора



Прогноз SARIMA



Прогноз Chronos



Прогноз Lag-Llama

## Результаты: Влияет ли частота данных на качество прогнозов LLM?

Значения ошибок RMSE для LLM приведены по отношению к RMSE прогнозов SARIMA, для учета специфики каждого ряда.

Модель	RMSE на тестовой выборке			
	Медиана	Медиана для годовых показателей	Медиана для квартальных показателей	Медиана для месячных показателей
Chronos	1.04	1.21	1.03	1.04
Lag-Llama	1.08	1.11	1.05	1.15



Гипотеза 2 о лучшей точности LLM на данных более высокой частоты **частично подтвердилась**. Наименьшей ошибкой в сравнении в SARIMA обладают прогнозы квартальных показателей. Возможные причины:

- LLM успешно применяет выученные паттерны к квартальным данным, в которых меньше шума и сезонности, чем в месячных, но больше контекста, чем в годовых.
- В связи с ограниченным числом данных годовой и квартальной частот доступных для российской экономики значения RMSE могут быть недостаточно репрезентативными, поскольку оценены на небольшом количестве тестовых точек.

Для годовых данных LLM применимы для прогнозирования тренда. Для квартальных и месячных данных LLM в целом применимы в одинаковой степени.

## Результаты: Влияет ли категория данных на качество прогнозов LLM?

Значения ошибок RMSE для LLM приведены по отношению к RMSE прогнозов SARIMA, для учета специфики каждого ряда.

Модель	Медиана	RMSE на тестовой выборке		
		Медиана для показателей реального сектора	Медиана для финансовых показателей	Медиана для опережающих показателей
Chronos	1.04	1.38	1.09	1.02
Lag-Llama	1.08	1.97	1.14	0.97

Гипотеза 3 о лучшей точности LLM для финансовых показателей **не подтвердилась**.

Точность прогнозов финансовых показателей выше, чем показателей реального сектора, но уступает прогнозам опережающих показателей. Возможные причины:

- Высокая точность прогнозов LLM на опережающих показателях связана со схожестью структуры этих данных – показателей операционной деятельности компаний и их ожидания относительно будущей активности – с теми, что были в обучении.

LLM слабо применимы к прогнозированию показателей реального сектора, но вполне справляются с финансовыми и опережающими индикаторами макроэкономики России.

## Заключение

1. LLM не имеет превосходства над SARIMA для прогнозирования макроэкономики России
2. LLM после предварительного сравнения с классическими эконометрическими моделями можно применять для прогнозирования квартальных и месячных финансовых и опережающих показателей экономики России, однако для годовых данных и реальных показателей предпочтительнее использовать статистические модели
3. Ключевая проблема применения уже обученных LLM для макроэкономического прогнозирования – недостаточность экономических данных в обучении моделей

## Дальнейшие исследования

1. Гибридные подходы, сочетающие гибкость и обобщающую способность LLM с теоретической строгостью эконометрических методов, например, включение структурированных экономических ограничений в архитектуру LLM
2. Способы обработки экономических данных для повышения качества прогнозов LLM, например, обеспечение стационарности данных



Национальный исследовательский  
университет «Высшая школа экономики»

Факультет  
экономических наук

Москва, 2025

# Оценка прогностического потенциала больших языковых моделей для наукастинга ключевых макроэкономических показателей России

Выполнила:

**Лашина Мария Дмитриевна**

Научный руководитель:

**Станкевич Иван Павлович**, к.э.н., доцент

Рецензент:

**Шибанов Олег Константинович**, к.ф.-м.н., PhD, профессор РЭШ



## Ответы на вопросы рецензента

1. Данные были разделены на обучающую и тестовую выборки. В процессе стандартизации среднее и отклонение были определены на обучающей выборке, затем применены к обоим выборкам.
2. Все модели прогнозируют одномерные временные ряды, т.е. модели были применены к каждому из 34 временных рядов. Для применения модели SARIMA порядок интегрирования для обеспечения стационарности был определен по критерию KSS. Для применения LLM стационарность не является обязательным условием, при этом данные в обучении не приводятся к стационарной форме. Вопрос о влиянии стационарности на качество прогноза LLM, можно рассмотреть в дальнейших исследованиях.
3. Все модели прогнозируют одномерные временные ряды. В данных представлены ряды с ежегодной, ежеквартальной и ежемесячной частотой. Модель SARIMA автоматически подбирает параметры и обучается на обучающей выборке, затем авторегрессионно прогнозирует наблюдения тестовой выборки, не обновляясь. LLM получают обучающую выборку в качестве входной последовательности, и также следующие наблюдения по авторегрессионной схеме.
4. Для показателя ИПЦ годовой частоты RMSE SARIMA 0.0569, RMSE Chronos 0.0041, их отношение составляет 0,0729, т.е. RMSE Chronos меньше в 13.7 раз, прогноз был сделан на год вперед для 4 лет.
5. В статье упомянутой рецензентом указано следующее: «No leakage can be ensured by using open-source LLMs with documented training data and published weights». Для исследования были взяты модели Chronos и Lag-Llama в открытом доступе, для которых указаны данные в обучении. Ни для одной из этих моделей в рамках обучения не использовались данные российской экономики, соответственно, утечки данных не было.
6. На тему оценки качества прогнозов LLM для экономических показателей все еще сделано мало исследований, однако, две работы упомянуты для сравнения результатов ((Bahelka & De Weerd, 2024), (Carriero, Pettenuzzo, & Shekhar, 2025)).

# Данные в обучении LLM: Chronos

URL: <https://arxiv.org/pdf/2403.07815>

Table 3: All datasets that are used for experiments. The datasets are partitioned according to how they are used for training and evaluation of CHRONOS models: *pretraining-only* data is only used for CHRONOS training; *in-domain evaluation* data is used for training CHRONOS models and other task-specific baselines, except for the *H* observations that are held out for in-domain testing only; *zero-shot evaluation* data is *not* used in training CHRONOS models, but only for evaluation (final *H* observations), as well as for training task-specific baselines (excluding the final *H* observations).

Dataset	Domain	Freq.	Num. Series	Series Length			Prediction
				min	avg	max	Length ( <i>H</i> )
Pretraining-only							
Brazilian Cities Temperature	nature	M	12	492	757	1320	
Mexico City Bikes	transport	1H	494	780	78313	104449	-
Solar (5 Min.)	energy	5min	5166	105120	105120	105120	-
Solar (Hourly)	energy	1H	5166	8760	8760	8760	-
Spanish Energy and Weather	energy	1H	66	35064	35064	35064	-
Taxi (Hourly)	transport	1H	2428	734	739	744	-
USHCN	nature	1D	6090	5906	38653	59283	-
Weatherbench (Daily)	nature	1D	225280	14609	14609	14610	-
Weatherbench (Hourly)	nature	1H	225280	350633	350639	350640	-
Weatherbench (Weekly)	nature	1W	225280	2087	2087	2087	-
Wiki Daily (100k)	web	1D	100000	2741	2741	2741	-
Wind Farms (Daily)	energy	1D	337	71	354	366	-
Wind Farms (Hourly)	energy	1H	337	1715	8514	8784	-
In-domain evaluation							
Electricity (15 Min.)	energy	15min	370	16032	113341	140256	24
Electricity (Hourly)	energy	1H	321	26304	26304	26304	24
Electricity (Weekly)	energy	1W	321	156	156	156	8
KDD Cup 2018	nature	1H	270	9504	10897	10920	48
London Smart Meters	energy	30min	5560	288	29951	39648	48
M4 (Daily)	various	1D	4227	107	2371	9933	14
M4 (Hourly)	various	1H	414	748	901	1008	48
M4 (Monthly)	various	1M	48000	60	234	2812	18
M4 (Weekly)	various	1W	359	93	1035	2610	13
Pedestrian Counts	transport	1H	66	576	47459	96424	48
Rideshare	transport	1H	2340	541	541	541	24
Taxi (30 Min.)	transport	30min	2428	1469	1478	1488	48

Taxi (30 Min.)	transport	30min	2428	1469	1478	1488	48
Temperature-Rain	nature	1D	32072	725	725	725	30
Uber TLC (Daily)	transport	1D	262	181	181	181	7
Uber TLC (Hourly)	transport	1H	262	4344	4344	4344	24
Zero-shot evaluation							
Australian Electricity	energy	30min	5	230736	231052	232272	48
CIF 2016	banking	1M	72	28	98	120	12
Car Parts	retail	1M	2674	51	51	51	12
Covid Deaths	healthcare	1D	266	212	212	212	30
Dominick	retail	1D	100014	201	296	399	8
ERCOT Load	energy	1H	8	154854	154854	154854	24
ETT (15 Min.)	energy	15min	14	69680	69680	69680	24
ETT (Hourly)	energy	1H	14	17420	17420	17420	24
Exchange Rate	finance	1B	8	7588	7588	7588	30
FRED-MD	economics	1M	107	728	728	728	12
Hospital	healthcare	1M	767	84	84	84	12
M1 (Monthly)	various	1M	617	48	90	150	18
M1 (Quarterly)	various	3M	203	18	48	114	8
M1 (Yearly)	various	1Y	181	15	24	58	6
M3 (Monthly)	various	1M	1428	66	117	144	18
M3 (Quarterly)	various	3M	756	24	48	72	8
M3 (Yearly)	various	1Y	645	20	28	47	6
M4 (Quarterly)	various	3M	24000	24	100	874	8
M4 (Yearly)	various	1Y	23000	19	37	841	6
M5	retail	1D	30490	124	1562	1969	28
NN5 (Daily)	finance	1D	111	791	791	791	56
NN5 (Weekly)	finance	1W	111	113	113	113	8
Tourism (Monthly)	various	1M	366	91	298	333	24
Tourism (Quarterly)	various	1Q	427	30	99	130	8
Tourism (Yearly)	various	1Y	518	11	24	47	4
Traffic	transport	1H	862	17544	17544	17544	24
Weather	nature	1D	3010	1332	14296	65981	30

# Данные в обучении LLM: Lag-Llama

URL: <https://arxiv.org/pdf/2310.08278>

**Table 3:** Datasets used in the pretraining corpus and the unseen datasets on which we evaluate, grouped by the domains they are labelled against.

	Transport & Tourism	Energy	Nature	Air Quality	Cloud	Banking & Econ
Pretraining	San Francisco Traffic Uber TLC Hourly	Australian Electricity Demand Electricity Hourly London Smart Meters Solar Wind Farms ETT H1 ETT H2 ETT M1	KDD Cup 2018 Sunspot	Beijing Multisite UCI	CPU Limit Minute CPU Usage Minute Function Delay Minute Instances Minute Memory Limit Minute Memory Usage Minute	
Unseen	Pedestrian Counts	ETT M2	Weather	Beijing PM2.5	Requests Minute Platform Delay Minute	Exchange Rate

## Список литературы (1/3)

1. Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., . . . al., e. (2023). Palm 2 technical report. Retrieved from [https://www.researchgate.net/publication/370842141\\_PaLM\\_2\\_Technical\\_Report#read](https://www.researchgate.net/publication/370842141_PaLM_2_Technical_Report#read)
2. Anonymous. (2023). Modeling time series as text sequence a frequency-vectorization transformer for time series forecasting. Submitted to The Twelfth International Conference on Learning Representations. Retrieved from <https://openreview.net/pdf?id=N1cjy5iznY>
3. Anonymous. (2023). Time series modeling at scale: a universal representation across tasks and domains. Submitted to The Twelfth International Conference on Learning Representations. Retrieved from <https://openreview.net/pdf?id=SZErAetdMu>
4. Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., . . . Rangapuram, S. S. (2024). Chronos: learning the language of time series. Transactions on Machine Learning Research. Retrieved from <https://arxiv.org/pdf/2403.07815>
5. Bahelka, A., & De Weerd, H. (2024). Comparative analysis of mixed-data sampling (MIDAS) model compared to Lag-Llama model for inflation nowcasting. Retrieved from <https://arxiv.org/html/2407.08510v1>
6. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Von Arx, S., . . . al., e. (2022). On the opportunities and risks of foundation models. Retrieved from <https://arxiv.org/pdf/2108.07258>
7. Borsos, Z., El Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., . . . al., e. (2023). Audioldm: a language modeling approach to audio generation. IEEE/ACM Transactions on Audio, Speech, and Language Processing. Retrieved from <https://arxiv.org/pdf/2209.03143>
8. Cai, Y., Goswami, M., Choudhry, A., Srinivasan, A., & Dubrawski, A. (2023). Jolt: jointly learned representations of language and time-series. Retrieved from <https://openreview.net/pdf?id=UVFIAMBj9u>
9. Cao, D., Jia, F., Arik, S. O., Pfister, T., Zheng, Y., Ye, W., & Liu, Y. (2023). Tempo: prompt based generative pre-trained transformer for time series forecasting. Retrieved from <https://arxiv.org/pdf/2310.04948>
10. Carriero, A., Pettenuzzo, D., & Shekhar, S. (2025). Macroeconomic forecasting with large language models. Retrieved from <https://arxiv.org/pdf/2407.00890>
11. Chang, C., Peng, W.-C., & Chen, T.-F. (2023). Llm4ts: two-stage fine-tuning for time-series forecasting with pretrained llms. Retrieved from <https://arxiv.org/pdf/2308.08469v2>
12. Chen, Y., Wang, X., & Xu, G. (2023). Gatgpt: a pre-trained large language model with graph attention network for spatiotemporal imputation. Retrieved from <https://arxiv.org/pdf/2311.14332>
13. Chung, H., Kim, J., Kwon, J.-m., Jeon, K.-H., Lee, M. S., & Choi, E. (2023). Text-to-ecg: 12 lead electrocardiogram synthesis conditioned on clinical text reports. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (pp. 1-5). Retrieved from <https://arxiv.org/pdf/2303.09395>
14. Ding, Y., Jia, S., Ma, T., Mao, B., Zhou, X., Li, L., & Han, D. (2023). Integrating stock features and global information via large language models for enhanced stock return prediction. Retrieved from <https://arxiv.org/pdf/2310.05627>
15. Duan, Y., Zhou, C., Wang, Z., Wang, Y.-K., & Lin, C.-t. (2023). Dewave: discrete encoding of eeg waves for eeg to text translation. Thirty-seventh Conference on Neural Information Processing Systems. Retrieved from <https://openreview.net/pdf?id=WaLI8slhLw>
16. Gao, H., Ni, J., Qian, K., Zhang, Y., Chang, S., & Hasegawa-Johnson, M. (2022). Wavprompt: towards few-shot spoken language understanding with frozen language models. Retrieved from <https://arxiv.org/pdf/2203.15863>
17. Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K. V., Joulin, A., & Misra, I. (2023). Imagebind: one embedding space to bind them all. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (pp. 15180–15190). Retrieved from <https://arxiv.org/pdf/2305.05665>
18. Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K. V., Joulin, A., & Misra, I. (2023). Imagebind: one embedding space to bind them all. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (pp. 15180–15190). Retrieved from <https://arxiv.org/pdf/2305.05665>
19. Gong, Y., Luo, H., Liu, A. H., Karlinsky, L., & Glass, J. (2023). Listen, think, and understand. Retrieved from <https://arxiv.org/pdf/2305.10790>
20. Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. (2023). Large language models are zero-shot time series forecasters. 37th Conference on Neural Information Processing Systems. Retrieved from <https://arxiv.org/pdf/2310.07820>
20. Han, W., Qiu, J., Zhu, J., Xu, M., Weber, D., Li, B., & Zhao, D. (2022). An empirical exploration of cross-domain alignment between language and electroencephalogram. Retrieved from <https://arxiv.org/pdf/2208.06348v3>

## Список литературы (2/3)

21. Han, W., Qiu, J., Zhu, J., Xu, M., Weber, D., Li, B., & Zhao, D. (2022). An empirical exploration of cross-domain alignment between language and electroencephalogram. Retrieved from <https://arxiv.org/pdf/2208.06348v3>
22. Hegselmann, S., Buendia, A., Lang, H., Agrawal, M., Jiang, X., & Sontag, D. (2023). Tabllm: few-shot classification of tabular data with large language model. International Conference on Artificial Intelligence and Statistics, (pp. 5549–5581). Retrieved from <https://arxiv.org/pdf/2210.10723v1>
23. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*, 9(8), 1735–1780. 30 Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., . . . al., e. (2023). Time-llm: time series forecasting by reprogramming large language models. Retrieved from <https://arxiv.org/pdf/2310.01728>
24. King, R., Yang, T., & Mortazavi, B. J. (2023). Multimodal pretraining of medical time series and notes. *Machine Learning for Health (ML4H)*, 244–255. Retrieved from <https://arxiv.org/pdf/2312.06855v1>
25. Kreuk, F., Synnaeve, G., Polyak, A., Singer, U., Defossez, A., Copet, J., . . . Adi, Y. (2022). Audiogen: textually guided audio. Retrieved from <https://arxiv.org/pdf/2209.15352>
26. Lakomkin, E., Wu, C., Fathullah, Y., Kalinli, O., Seltzer, M. L., & Fuegen, C. (2023). End-to end speech recognition contextualization with large language models. Retrieved from <https://arxiv.org/pdf/2309.10917>
27. Leng, Z., Kwon, H., & Plotz, T. (2023). Generating virtual on-body accelerometer data from virtual textual descriptions for human activity recognition. Retrieved from <https://arxiv.org/pdf/2305.03187>
28. Liu, C., Wan, Z., Cheng, S., Zhang, M., & Arcucci, R. (2023). Etp: learning transferable ecg representations via ecg-text pre-training. Retrieved from <https://arxiv.org/pdf/2309.07145>
29. Liu, C., Yang, S., Xu, Q., Li, Z., Long, C., Li, Z., & Zhao, R. (2024). Spatial-temporal large language model for traffic prediction. Retrieved from <https://arxiv.org/pdf/2401.10134v1>
30. Liu, S., Hussain, A. S., Sun, C., & Shan, Y. (2023). Music understanding llama: advancing text to-music generation with question answering and captioning. Retrieved from <https://arxiv.org/pdf/2308.11276>
31. Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., & Zimmermann, R. (2023). Unitime: a language-empowered unified model for cross-domain time series forecasting. Retrieved from <https://arxiv.org/pdf/2310.09751>
32. Liu, X., McDuff, D., Kovacs, G., Galatzer-Levi, I., Sunshine, J., Zhan, J., . . . Patel, S. (2023). Large language models are few-shot health learners. Retrieved from <https://arxiv.org/pdf/2305.15525>
33. Mirchandani, S., Xia, F., Florence, P., Ichter, B., Driess, D., Arenas, M. G., . . . Zeng, A. (2023). Large language models as general pattern machines. Retrieved from <https://arxiv.org/pdf/2307.04721>
34. Moon, S., Madotto, A., Lin, Z., Dirafzoon, A., Saraf, A., Bearman, A., & Damavandi, B. (2022). Imu2clip: multimodal contrastive learning for imu motion sensors from egocentric videos and text. Retrieved from <https://arxiv.org/pdf/2210.14395v1>
35. Qin, Y., Liang, S., Ye, Y., Zhu, K., Yan, L., Lu, Y., . . . al., e. (2023). Toolllm: facilitating large language models to master 16000+ real-world apis. Retrieved from <https://arxiv.org/pdf/2307.16789>
36. Qiu, J., Han, W., Zhu, J., Xu, M., Rosenberg, M., Liu, E., . . . Zhao, D. (2023). Llama 2: open foundation and fine-tuned chat models. Retrieved from <https://arxiv.org/pdf/2307.09288>
37. Rasul, K., Ashok, A., Williams, A. R., Ghonia, H., Bhagwatkar, R., Khorasani, A., & al., e. (2024). Lag-llama: towards foundation models for probabilistic time series forecasting. Retrieved from <https://arxiv.org/pdf/2310.08278>
38. Rubenstein, P. K., Asawaroengchai, C., Nguyen, D. D., Bapna, A., Borsos, Z., De Chaumont-Quitry, F., . . . al., e. (2023). Audiopalm: a large language model that can speak and listen. Retrieved from <https://arxiv.org/pdf/2306.12925>
39. Schuster, M. a. (1997). Bidirectional neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673–2681. doi:10.1109/78.650093
40. Su, Y., Lan, T., Li, H., Xu, J., Wang, Y., & Cai, D. (2023). Pandagpt: one model to instruction follow them all. Retrieved from <https://arxiv.org/pdf/2305.16355>



## Список литературы (<sup>3</sup>/<sub>3</sub>)

41. Sun, C., Li, Y., Li, H., & Hong, S. (2023). Test: text prototype aligned embedding to activate llm's ability for time series. Retrieved from <https://arxiv.org/pdf/2308.08241>
42. Tang, C., Yu, W., Sun, G., Chen, X., Tan, T., Li, W., . . . Zhang, C. (2023). Salmonn: towards generic hearing abilities for large language models. Retrieved from <https://arxiv.org/pdf/2310.13289>
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. Retrieved from <https://arxiv.org/abs/1706.03762>
44. Wang, T., Zhou, L., Zhang, Z., Wu, Y., Liu, S., Gaur, Y., . . . Wei, F. (2023). Viola: unified codec language models for speech recognition, synthesis, and translation. Retrieved from <https://arxiv.org/pdf/2305.16107>
45. Wang, X., Fang, M., Zeng, Z., & Cheng, T. (2023). Where would I go next? Large language models as human mobility predictors. Retrieved from <https://arxiv.org/pdf/2308.15197v1>
46. Wang, Z., & Ji, H. (2022). Open vocabulary electroencephalography-to-text decoding and zero shot sentiment classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, pp. 5350–5358. Retrieved from <https://arxiv.org/pdf/2112.02690>
47. Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). Transformers in time series: a survey. Retrieved from <https://arxiv.org/pdf/2202.07125>
48. Wimmer, C., & Rekabsaz, N. (2023). Leveraging vision-language models for granular market change prediction. Retrieved from <https://arxiv.org/pdf/2301.10166>
49. Wu, S., Irsoy, O., Lu, S., Dabrowski, V., Dredze, M., Gehrmann, S., . . . Mann, G. (2023). Bloomberggpt: a large language model for finance. Retrieved from <https://arxiv.org/pdf/2303.17564>
50. Xie, Q., Han, W., Lai, Y., Peng, M., & Huang, J. (2023). The wall street neophyte: a zero-shot analysis of chatgpt over multimodal stock movement prediction. Retrieved from <https://arxiv.org/pdf/2304.05351>
51. Xue, H., & Salim, F. D. (2023). PromptCast: A New Prompt-based Learning Paradigm for Time Series Forecasting. Retrieved from <https://arxiv.org/pdf/2210.08964>
52. Xue, H., & Salim, F. D. (2023). Utilizing language models for energy load forecasting. Retrieved from <https://arxiv.org/pdf/2310.17788>
53. Xue, H., Voutharoja, B. P., & Salim, F. D. (2022). Leveraging Language Foundation Models for Human Mobility Forecasting. Retrieved from <https://arxiv.org/pdf/2209.05479>
54. Yang, D., Tian, J., Tan, X., Huang, R., Liu, S., Chang, X., . . . al., e. (2023). Uniaudio: an audio foundation model toward universal audio generation. Retrieved from <https://arxiv.org/pdf/2310.00704>
55. Yu, X., Chen, Z., Ling, Y., Dong, S., Liu, Z., & Lu, Y. (2023). Temporal data meets llm explainable financial time series forecasting. Retrieved from <https://arxiv.org/pdf/2306.11025>
56. Zhang, X., Chowdhury, R. R., Gupta, R. K., & Shang, J. (2025). Large language models for time series: a survey. *ACM Transactions on Intelligent Systems and Technology*, 16(3), 1–20. Retrieved from <https://arxiv.org/pdf/2308.08469>
57. Zhang, Y., Zhang, Y., Zheng, M., Chen, K., Gao, C., Ge, R., . . . al., e. (2023). Insight miner: a time series analysis dataset for cross-domain alignment with natural language. *NeurIPS 2023 AI for Science Workshop*. Retrieved from <https://openreview.net/pdf?id=E1khscdUdH>
58. Zhang, Z., Amiri, H., Liu, Z., & Zufle, A. (2023). Large language models for spatial trajectory patterns mining. Retrieved from <https://arxiv.org/pdf/2310.04942>
59. Zhong, Z., Rempe, D., Chen, Y., Ivanovic, B., Cao, Y., Xu, D., . . . Ray, B. (2023). Language guided traffic simulation via scene-level diffusion. Retrieved from <https://arxiv.org/pdf/2306.06344>
60. Zhou, T., Niu, P., Wang, X., Sun, L., & Rong, J. (2023). One fits all: power general time series analysis by pretrained lm. Retrieved from <https://arxiv.org/pdf/2302.11939>
61. Zhou, Y., Yang, J., Zou, H., & Xie, L. (2023). Tent: connect language models with iot sensors for zero-shot activity recognition. Retrieved from <https://arxiv.org/pdf/2311.08245>