Банк России

ДИСКУССИЯ К РАБОТЕ «КАК БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ ПОМОГАЮТ В ПОСТРОЕНИИ ОПЕРЕЖАЮЩИХ ИНДИКАТОРОВ МАКРОЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ?»

Сергей Селезнев

Департамент исследований и прогнозирования
Банк России

30 июня, 2026
СОВМЕСТНЫЙ СЕМИНАР БАНКА РОССИИ, РЭШ И НИУ ВШЭ
«ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЭКОНОМИКА», САНКТ-ПЕТЕРБУРГ

О чем эта работа?

Идея

- Новости – дополнительный источник неструктурированных данных, который содержит дополнительную к «твердым» данным информацию. Эту информацию можно попытаться использовать для улучшения качества экономических прогнозов.

Особенности работы

- LLM (большая языковая модель) + специализированная процедура агрегирования новостей в индекс на основе RAG, саммаризации и релевантности;
- Российские данные.

Практическое приложение и выводы

- Прогнозирование индекса промышленного производства и отраслевых индексов;
- Добавление новостного индекса улучшает качество прогноза модели.

Сильные стороны работы

Тема очень актуальна

- Развитие LLM открывает перед исследователями новые возможности для извлечения информации из текста за счет увеличения качества проводимого анализа по сравнению с применяемыми ранее методами (*Ash and Hansen (2023)*), а также снижения трудозатрат и стоимости на извлечение информации (*Asirvatham et al. (2026)*);
- Исследователи по всему миру пытаются адаптировать эти преимущества для повышения качества прогнозов экономических показателей (*Kwon et al. (2025)*, *Lopez-Lira and Tang (2025)*, …);
- Одна из первых подобного рода работ для России.

Расширенный алгоритм построения индекса

- Рассуждения + RAG + саммаризация + релевантность – комбинация, которая расширяет классическую парадигму построения новостных индексов на основе LLM;
- Понятная мотивация для использования LLM и отдельных модификаций стандартной процедуры разметки (рассуждения, RAG, саммаризация, релевантность) с наглядными иллюстрациями.

Вопросы по дизайну экспериментов

Выигрыш в качестве прогнозов на тестовом периоде в 1 год – статистическая закономерность или специфичный для конкретного периода/фазы бизнес-цикла результат?

- Значимо ли статистически различие в качестве прогнозов с новостным индексом и без?
- Насколько репрезентативна тестовая выборка длиной в 1 год?

Сохраняется ли выигрыш в прогнозе для более классических моделей?

- Стекинг случайного леса и векторной авторегрессии – сложный и малоизученный со статистической точки зрения инструмент. Каковы его свойства при добавлении неинформативного показателя в список предикторов? («шумность» - один из ингредиентов качества работы случайного леса);
- Дают ли линейные регрессионные методы выигрыш при добавлении новостей?

Нужна ли такая сложная система?

Нужно ли использовать LLM?

- LLM – достаточно вычислительно сложная и относительно более дорогая по сравнению с классическими моделями альтернатива;
- Как ведут себя модели с индексами на основе более простых подходов?

Каждый из компонентов системы дает вклад в увеличение прогнозной точности?

- Обычно исследователи в экономике следуют простой стратегии: оценил тональность новостей → усреднил результаты.
- Данная работа добавляет как минимум три не совсем стандартные модификации к этому методу: RAG, саммаризация, релевантность.
- Что дает каждая из них не с точки зрения гладкости и шумности индекса, а с точки зрения его прогнозных свойств?

Внутривыборочные смещения – главный (открытый) вопрос при работе с временными рядами на основе LLM

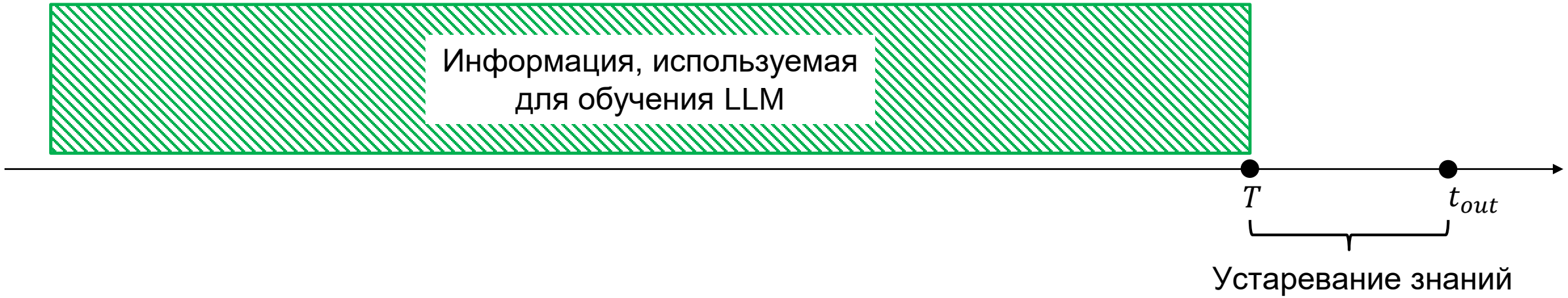
Статистические свойства индекса внутри выборки обучения LLM и вне ее могут быть разными

- Индекс – сложная функция от новостей и весов LLM (обучающей выборки LLM):

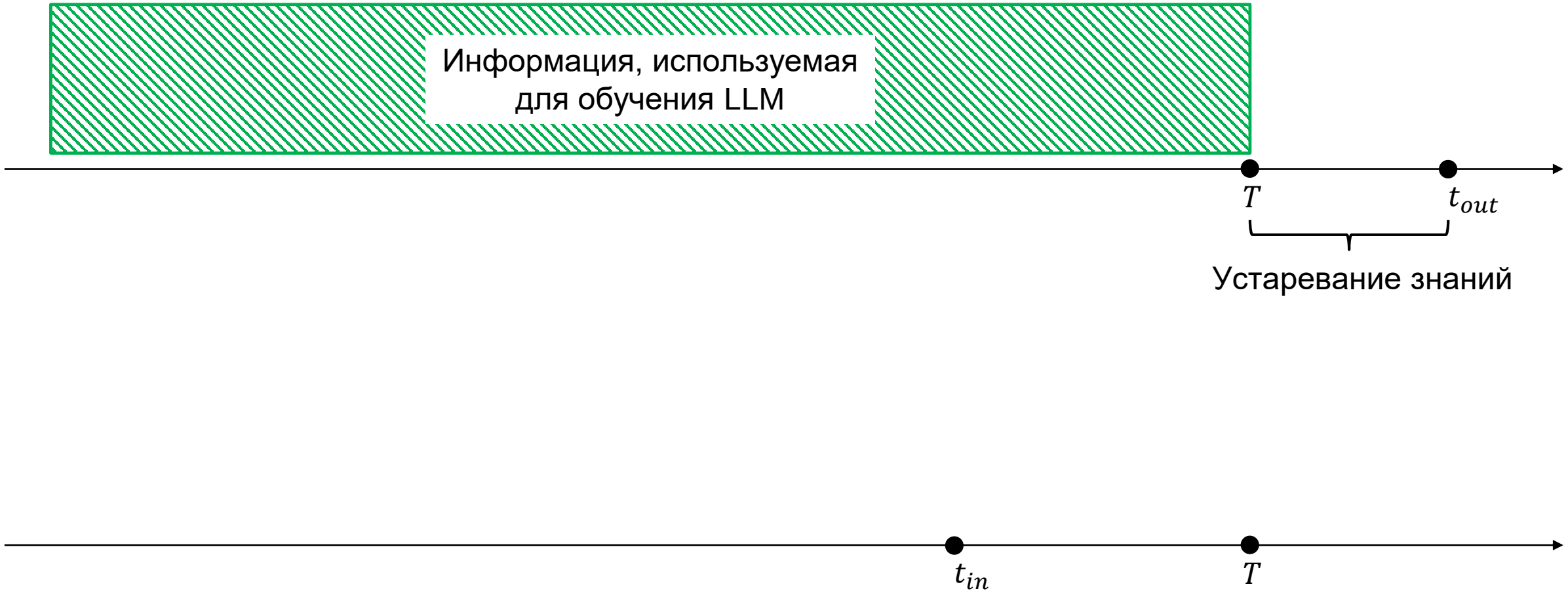
$$I_{t_{out}}^{\text{вне выборки}} = f\left(\begin{matrix}\text{новости}_{:t_{out}} \\ \text{информация до границы отсечения модели } T\end{matrix}\right)$$

$$I_{t_{in}}^{\text{внутри выборки}} = f\left(\begin{matrix}\text{новости}_{:t_{in}} \\ \text{информация до } t_{in} - (t_{out} - T) \\ \text{информация от } t_{in} - (t_{out} - T) \text{ до } t_{in} \\ \text{информация от } t_{in} \text{ до } T\end{matrix}\right)$$

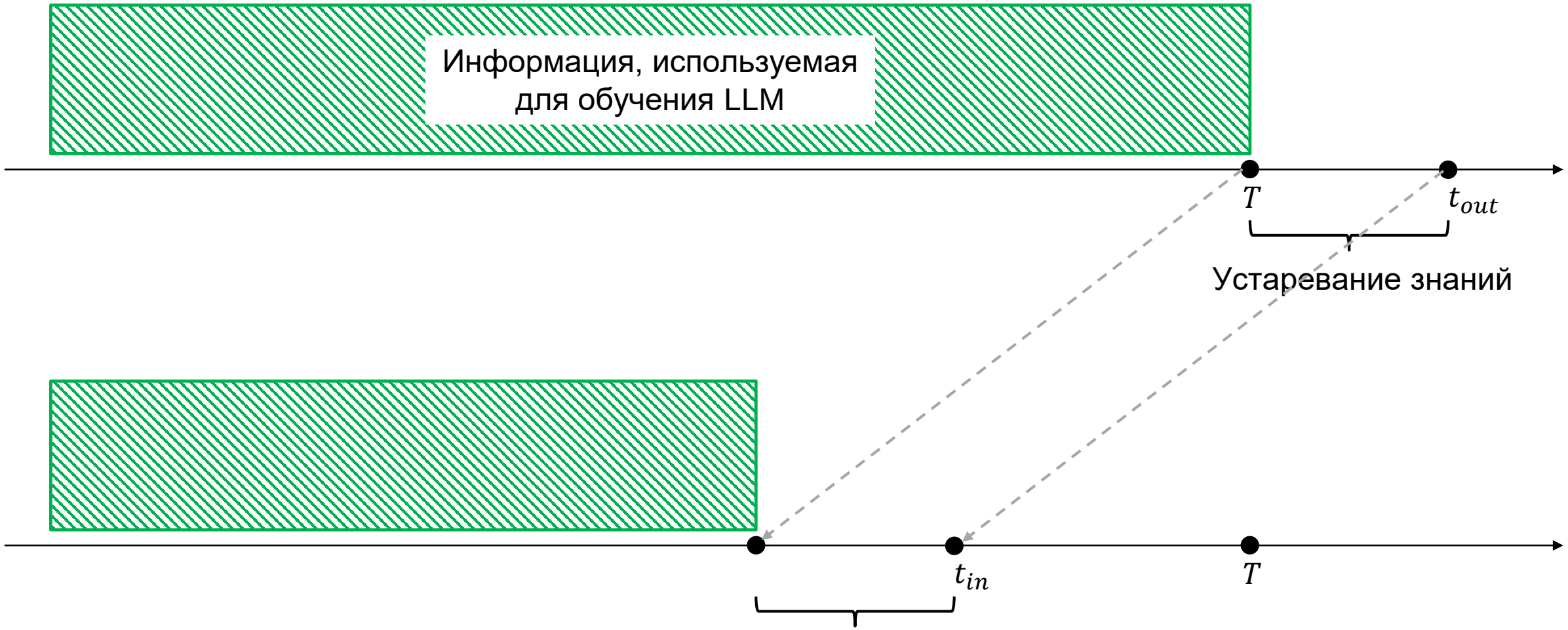
Тайминг в LLM



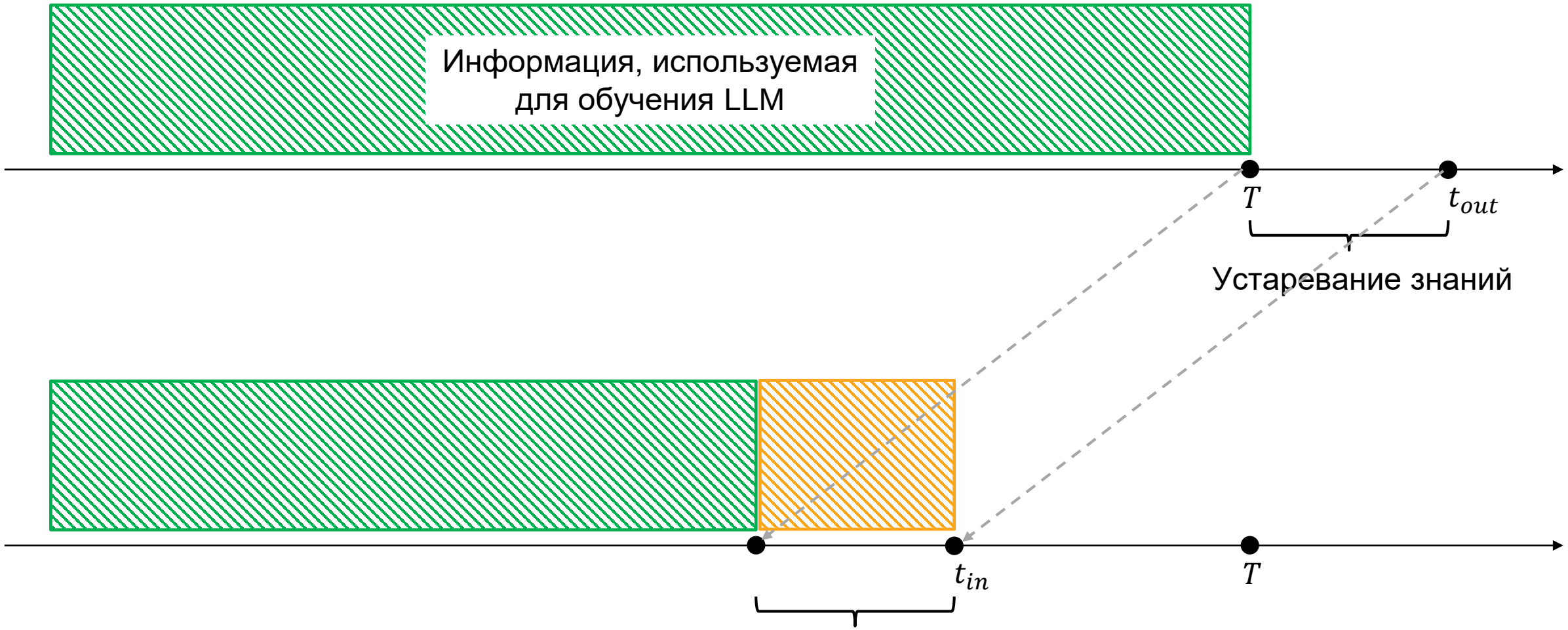
Тайминг в LLM



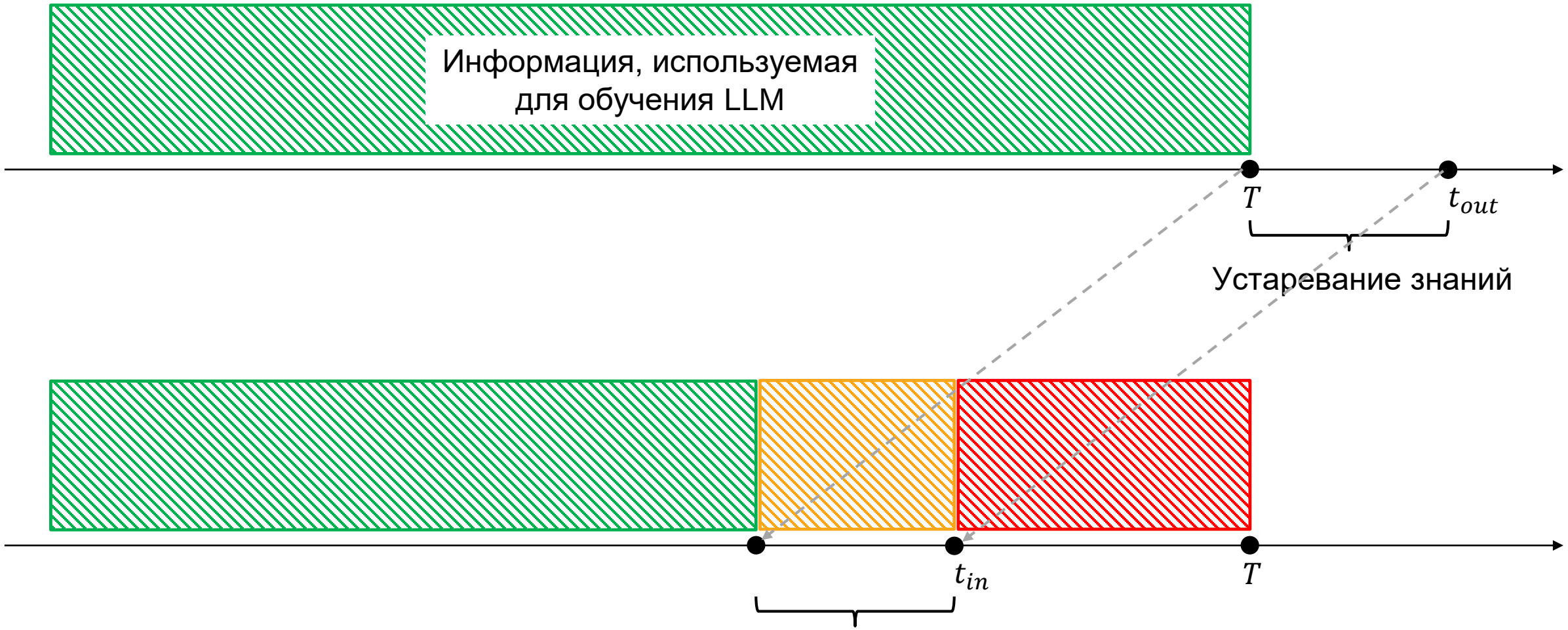
Тайминг в LLM



Тайминг в LLM



Тайминг в LLM



Какие потенциальные проблемы это создает?

Смещенные оценки новостного индекса внутри выборки

- Индекс внутри выборки имеет другие статистические свойства по сравнению с индексом вне выборки обучения LLM.

Смещение концепции (Concept shift)

- Связь прогнозируемой переменной и предиктора внутри и вне выборки обучения LLM становятся разными (*Murphy (2023)*).

Тестирование индекса внутри выборки обучения LLM не отражает свойств вне

- Цель тестирования – понять как модель с индексом будет вести себя в будущем (вне периода обучения LLM);
- Точки внутри выборки как минимум не могут быть выбраны для тестирования модели.

Что существующая литература говорит по этому поводу?

Проблема по крайней мере для некоторых задач прогнозирования реально существует

- *Sarkar and Vafa (2024*): прогнозирование рисков компаний и результатов выборов;
- *Alam et al. (2026), Eliseev and Seleznev (2026)*: макроэкономическое прогнозирование.

Эконометрика поверх результатов из LLM

- *Ludwig et al (2026)*: условия при которых измерения на основе LLM можно использовать в прогнозировании и оценке причинно-следственных связей.

Возможно проблема не так велика в простых задачах разметки

- *Asirvatham et al. (2026)*: хорошее описание проблемы смещений в LLM и ряд проверок для широкого круга задач разметки, в которых ChatGPT не показывают признаков смещения.

Некоторые пути контроля и избегания проблем со смещениями

Тесты на возможные смещения

- *Asirvatham et al. (2026)*: подбирать под задачи тесты, которые могут показать отсутствие смещений, а при наличии корректировать возникающее смещение.

Корректировка задачи таким образом, чтобы смещения не возникали

- *Engelberg et al. (2025)*: модифицировать задачу путем «анонимизации» новости, чтобы LLM не могла понять ее времени и объектов, максимально сохраняя при этом экономический смысл.

Использовать модели в которых нет «утечек»

- *He et al. (2025a, 2025b)*: семейство моделей, которые обучены до определенной даты для контроля проблемы «заглядывания вперед» (строят модели знания которых ограничены определенным годом).

Список литературы (1)

- *Alam, M.J., Boyle, S., Li, H. and Sekhposyan, T. (2026). ChatMacro: Evaluating Inflation Forecasts of Generative AI. Federal Reserve Bank of San Francisco Working Paper 2026-04.*
- *Ash, E. and Hansen, S. (2023). Text Algorithms in Economics. Annual Review Economics. 15:659-688.*
- *Asirvatham, H., Mokski, E. and Shleifer, A. (2026). GPT as a Measurement Tool. NBER Working Paper 34834.*
- *Engelberg, J., Manela, A., Mullins, W. and Vulicevic, L. (2025). Entity Neutering. Available at SSRN.*
- *Eliseev, A. and Seleznev, S. (2026). Fake date tests: Can We Trust In-Sample Accuracy of LLMs in Macroeconomic Forecasting? Bank of Russia Working Paper Series 167.*
- *He, S., Lv, L., Manela, A. and Wu, J. (2025a). Chronologically Consistent Large Language Models. arXiv:2502.21206.*

Список литературы (2)

- *He, S., Lv, L., Manela, A. and Wu, J. (2025b). Instruction Tuning Chronologically Consistent Language Models. arXiv:2510.11677.*
- *Kwon, B., Park, T., Rungcharoenkitkul, P. and Smets, F. (2025). Parsing the Pulse: Decomposing Macroeconomic Sentiment with LLMs. BIS Working Papers 1294.*
- *Lopez-Lira, A. and Tang, Y. (2025). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. arXiv:2304.07619.*
- *Ludwig, J., Mullainathan, S. and Rambachan, A. (2026). Large Language Models: An Applied Econometric Framework. Annual Review Economics. 18:In press.*
- *Murphy, K.P. (2023). Probabilistic Machine Learning: Advanced Topics. MIT Press.*
- *Sarkar, S., and Vafa, K. (2024). Lookahead Bias in Pretrained Language Models. Available at SSRN.*