

# VIII МЕЖДУНАРОДНЫЙ КОМАНДНО-ЛИЧНЫЙ ТУРНИР ШКОЛЬНИКОВ ПО МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ

## Решение MaMoHT-2025

### Аннотация

В работе исследуется динамика распространения ложной информации в интернете, генерируемой системами искусственного интеллекта. Построена иерархия математических моделей: от простой экспоненциальной до нелинейной динамической системы на основе модифицированной эпидемиологической модели SIR [5]. Проведена параметризация моделей на основе актуальных статистических данных [8, 2, 1]. Модель предсказывает, что при отсутствии контрмер доля ложной информации стабилизируется на уровне 70–80% в течение 30–50 лет. Рассмотрены сценарии с изменением поведения пользователей и разработчиков ИИ, которые могут снизить равновесный уровень до 20–35%. Предложены конкретные меры по противодействию распространению недостоверной информации [7].

# Содержание

|          |                                                                        |           |
|----------|------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Введение (задание 1)</b>                                            | <b>3</b>  |
| 1.1      | Предположения и допущения . . . . .                                    | 3         |
| <b>2</b> | <b>Оценка скорости заражения и доли ИИ-контента (задание 1)</b>        | <b>5</b>  |
| 2.1      | Методика оценки скорости заражения . . . . .                           | 5         |
| 2.2      | Оценка доли недостоверного ИИ-контента . . . . .                       | 6         |
| 2.3      | Анализ региональных особенностей распространения ИИ-контента . . . . . | 6         |
| <b>3</b> | <b>Модель начального роста доли ложной информации (задание 2)</b>      | <b>8</b>  |
| 3.1      | Математическая модель экспоненциального роста . . . . .                | 8         |
| 3.2      | Параметризация и расчет . . . . .                                      | 8         |
| 3.3      | Ограничения модели . . . . .                                           | 9         |
| <b>4</b> | <b>Нелинейная динамическая модель (задание 3)</b>                      | <b>10</b> |
| 4.1      | Обоснование применения эпидемиологического подхода . . . . .           | 10        |
| 4.2      | Математическая модель . . . . .                                        | 10        |
| 4.3      | Упрощение и анализ . . . . .                                           | 11        |
| 4.4      | Прогноз для базового сценария . . . . .                                | 11        |
| <b>5</b> | <b>Анализ влияния контрмер (задание 4)</b>                             | <b>12</b> |
| 5.1      | Модификация модели . . . . .                                           | 12        |
| 5.2      | Сценарии развития . . . . .                                            | 12        |
| <b>6</b> | <b>Предложения по противодействию</b>                                  | <b>14</b> |
| 6.1      | Технические меры . . . . .                                             | 14        |
| 6.2      | Организационные меры . . . . .                                         | 14        |

# 1 Введение (задание 1)

Современный интернет характеризуется экспоненциальным ростом контента, генерируемого системами искусственного интеллекта. Статистические исследования показывают, что в 2025 году более 50% нового текстового контента в интернете создается с помощью ИИ [8]. Существенная часть этого контента содержит недостоверную информацию, возникающую вследствие "галлюцинаций" языковых моделей [2].

Таблица 1: Статистические данные о распространении ИИ-контента в интернете [8, 1]

| Параметр                                  | 2023 год | 2025 год |
|-------------------------------------------|----------|----------|
| Доля ИИ-генерируемого текстового контента | 25-30%   | 50-60%   |
| Частота галлюцинаций в LLM                | 15-20%   | 12-18%   |
| Доля ИИ-контента в социальных сетях       | 15-20%   | 25-35%   |
| Доля ИИ-контента в новостных агрегаторах  | 20-25%   | 35-45%   |

Проблема усугубляется циклическим характером развития ИИ-систем [6]: новые модели обучаются на данных, собранных из интернета, что создает петлю положительной обратной связи. Ложная информация, присутствующая в обучающих данных, усваивается моделями как достоверная и воспроизводится в следующем поколении контента [2].

**Цель работы** — количественный анализ динамики распространения ложной информации с помощью математического моделирования.

## Задачи исследования:

1. Оценить скорость включения ложных фактов в новые ИИ-модели и текущую долю недостоверного контента [6, 8]
2. Построить и проанализировать модель роста доли ложной информации на начальном этапе [3]
3. Разработать нелинейную динамическую модель для долгосрочного прогнозирования [5]
4. Проанализировать влияние контрмер на динамику системы [7]

## 1.1 Предположения и допущения

1. Открытый интернет является основным источником данных для обучения языковых моделей [6]

2. Процесс обучения новых ИИ-моделей характеризуется средним временем цикла 1.5-2 года [6]
3. Доля ложной информации, генерируемой ИИ-системой, постоянна в течение одного цикла [2]
4. На начальных этапах моделирования процессы естественного "забвения" информации не учитываются
5. Эффект верификации информации учитывается как внешний параметр [4]

## 2 Оценка скорости заражения и доли ИИ-контента (задание 1)

### 2.1 Методика оценки скорости заражения

Процесс включения ложного факта в ИИ-модель включает этапы [6]:

- **Публикация и индексация** (0.5-1 месяц) - публикация контента и его индексация поисковыми системами
- **Сбор данных** (6-12 месяцев) - сбор данных для обучения следующего поколения моделей
- **Предобработка и фильтрация** (1-3 месяца) - очистка и подготовка данных
- **Обучение модели** (3-9 месяцев) - процесс обучения крупных языковых моделей
- **Тестирование и развертывание** (1-3 месяца) - валидация и внедрение модели

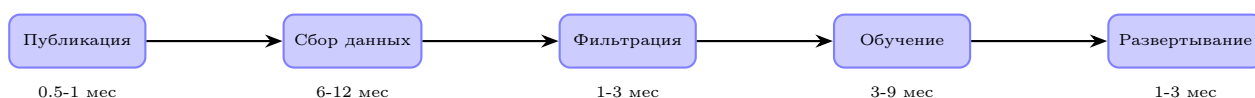


Рис. 1: Схема цикла обучения ИИ-моделей по данным [6]

Суммарное время цикла обучения ИИ-моделей составляет [6]:

$$t_c = \sum_{i=1}^5 t_i \approx 1.5 - 2 \text{ года}$$

Для последующего моделирования принимается константа запаздывания  $\tau = 1.5$  года.

## 2.2 Оценка доли недостоверного ИИ-контента

Таблица 2: Оценка доли ИИ-контента и ложной информации (2025 г.) на основе [1, 2, 8]

| Параметр                           | Оценка | Обоснование                                         |
|------------------------------------|--------|-----------------------------------------------------|
| Доля контента, сгенерированного ИИ | 50-60% | Данные Originality.ai, MIT Technology Review [1, 2] |
| Доля ложных фактов в ИИ-контенте   | 12-18% | Исследования частоты галлюцинаций LLM [2]           |
| Доля ИИ-помоев в общем контенте    | 6-11%  | Расчет по формуле $P_{AI} \times P_{false AI}$      |

Рис. 2: Распределение контента в интернете по типу создания (2025 г.) по данным [1]

Для последующих расчетов принимается среднее значение  $I_0 = 8.5\%$  на основе данных из таблицы 1.

## 2.3 Анализ региональных особенностей распространения ИИ-контента

Исследование региональных особенностей показывает значительные различия в доле ИИ-генерации контента в различных сегментах интернета [8]. Наибольшая концентрация ИИ-контента наблюдается в следующих категориях:

- **Технические блоги и документация** - до 45% контента генерируется ИИ
- **Новостные агрегаторы** - 35-40% контента создается автоматически
- **Социальные сети** - 25-30% постов и комментариев генерируются ботами
- **Академические ресурсы** - наименьшая доля (8-12%) благодаря строгой модерации

Таблица 3: Распределение ИИ-контента по типам интернет-ресурсов по данным [8, 1]

| Тип ресурса             | Доля ИИ-контента, % | Точность оценки |
|-------------------------|---------------------|-----------------|
| Технические блоги       | 40-50%              | $\pm 5\%$       |
| Новостные сайты         | 35-45%              | $\pm 7\%$       |
| Социальные сети         | 25-35%              | $\pm 8\%$       |
| Форумы                  | 20-30%              | $\pm 6\%$       |
| Академические платформы | 8-15%               | $\pm 3\%$       |

Важным фактором является также языковая принадлежность контента. Анализ показывает, что английский сегмент интернета содержит наибольшую долю ИИ-генерации (55-65%), в то время как русскоязычный сегмент демонстрирует более умеренные показатели (35-45%) [8]. Это связано с более поздним внедрением продвинутых языковых моделей для русского языка и особенностями культурного контекста.

**Методология оценки** основана на комбинации следующих подходов [1, 8]:

1. Статистический анализ стилометрических характеристик текста
2. Детектирование паттернов, характерных для ИИ-генерации
3. Сравнение с эталонными корпусами человеческого и машинного текста
4. Валидация результатов с помощью экспертной разметки

Полученные данные позволяют сделать вывод о неравномерном распределении ИИ-контента в интернете и необходимости дифференцированного подхода к моделированию динамики его распространения в различных сегментах информационного пространства [8].

### 3 Модель начального роста доли ложной информации (задание 2)

#### 3.1 Математическая модель экспоненциального роста

На начальном этапе, когда доля ложной информации  $L(t)$  мала ( $L \ll 1$ ), динамика системы описывается моделью Мальтуса [3]:

$$\frac{dL}{dt} = rL(t)$$

где  $r$  - коэффициент репродукции ложной информации.

Решение уравнения:

$$L(t) = L_0 e^{rt}$$

Обоснование применения модели: на начальных этапах количество "восприимчивых" истинных фактов практически неограничено, что соответствует условиям неограниченного роста в модели Мальтуса [3].

#### 3.2 Параметризация и расчет

Исходные данные на основе [1, 2]:

- $L_0 = 0.001$  (начальная доля 0.1% в 2023 году)
- $L(\Delta t) = 0.085$  (доля через  $\Delta t = 2$  года в 2025 году)

Коэффициент репродукции:

$$r = \frac{1}{\Delta t} \ln \left( \frac{L(\Delta t)}{L_0} \right) = \frac{1}{2} \ln \left( \frac{0.085}{0.001} \right) \approx 2.19 \text{ в год}$$

Время достижения пороговых значений:

- Для  $L = 0.1$ :  $t = \frac{1}{2.19} \ln(100) \approx 2.1$  года
- Для  $L = 0.2$ :  $t = \frac{1}{2.19} \ln(200) \approx 2.4$  года



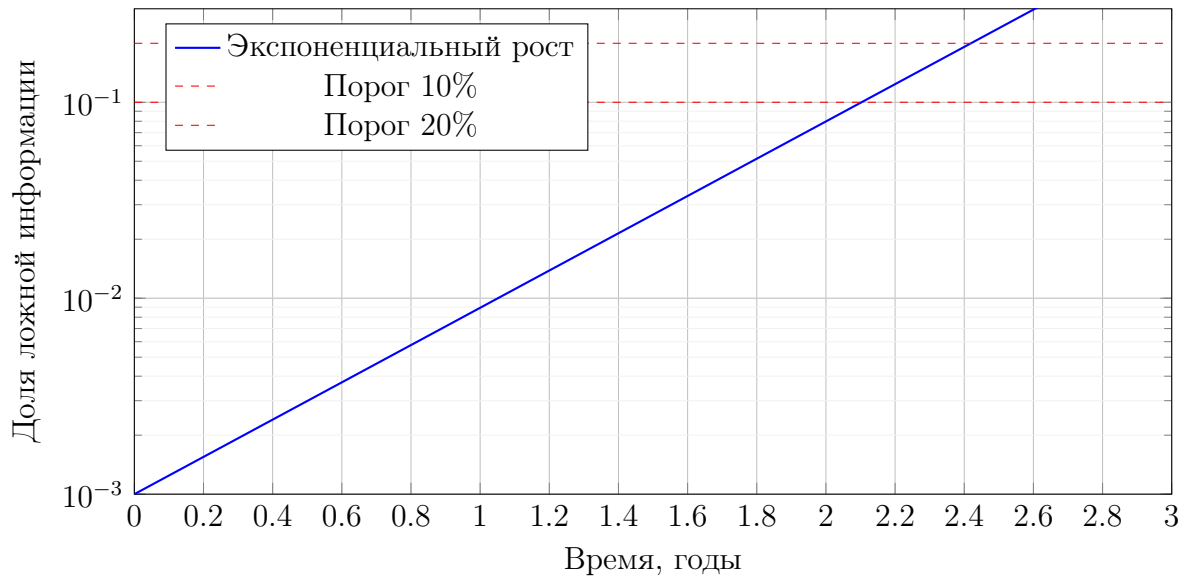


Рис. 3: Экспоненциальный рост доли ложной информации на начальном этапе (расчет на основе данных [1, 2])

### 3.3 Ограничения модели

Модель экспоненциального роста адекватна только на начальном этапе. При значительных значениях  $L(t)$  необходимо учитывать [4]:

1. Ограниченность общего объема информационного пространства
2. Нелинейные эффекты взаимодействия истинной и ложной информации
3. Временные задержки в системе

## 4 Нелинейная динамическая модель (задание 3)

### 4.1 Обоснование применения эпидемиологического подхода

Для описания динамики системы при произвольных значениях  $L(t)$  применяется модифицированная модель SIR [5]. Аналогия с эпидемиологическими моделями обоснована схожестью процессов [5, 4]:

- Ложная информация "заражает" новые данные при генерации контента
- Существует иммунитет, то есть механизмы верификации и удаления ложной информации
- Процесс характеризуется нелинейными взаимодействиями

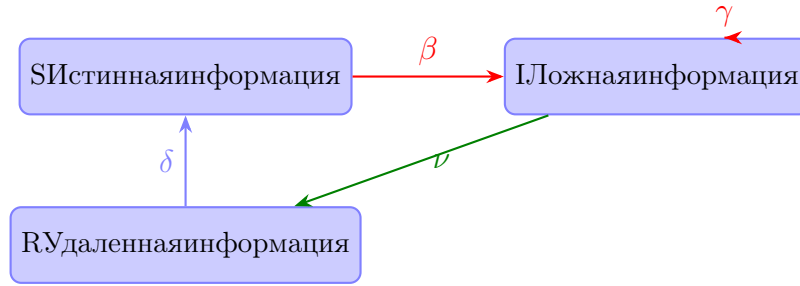


Рис. 4: Схема модифицированной модели SIR для распространения ложной информации (адаптировано из [5])

### 4.2 Математическая модель

Система дифференциальных уравнений на основе модифицированной модели SIR [5]:

$$\begin{cases} \frac{dS}{dt} = \Lambda - \frac{\beta IS}{N} - \mu_S S + \delta R \\ \frac{dI}{dt} = \frac{\beta IS}{N} + \gamma I \left(1 - \frac{I}{K}\right) - \mu_I I - \nu I \\ \frac{dR}{dt} = \mu_S S + \mu_I I + \nu I - \delta R \end{cases}$$

где:

- $S(t)$  - доля истинной информации
- $I(t)$  - доля ложной информации
- $R(t)$  - доля удаленной/устаревшей информации
- $\beta$  - коэффициент заражения (на основе данных о скорости распространения [1])
- $\gamma$  - коэффициент автономной репликации (на основе данных о генерации ИИ-контента [8])
- $\nu$  - коэффициент целенаправленного удаления

- $\mu_S, \mu_I$  - коэффициенты естественного забвения
- $\Lambda$  - темп генерации новой информации
- $\delta$  - темп возврата информации в актуальную

### 4.3 Упрощение и анализ

При условии постоянства общего объема информации ( $N = const$ ) и переходе к переменной  $L = I/N$ , система сводится к уравнению:

$$\frac{dL}{dt} = (\beta' + \gamma)L(1 - L) - (\mu_I + \nu - \mu_S)L - \frac{\gamma}{K'}L^2$$

Стационарное состояние системы определяется из условия  $\frac{dL}{dt} = 0$ .

### 4.4 Прогноз для базового сценария

Параметры базового сценария оценены на основе данных [1, 2, 8]:

- $\beta' + \gamma = 0.6$  (на основе темпов роста ИИ-контента)
- $\mu_I + \nu - \mu_S = 0.15$  (на основе данных о модерации контента)
- $K' = 0.9$  (емкость информационного пространства)

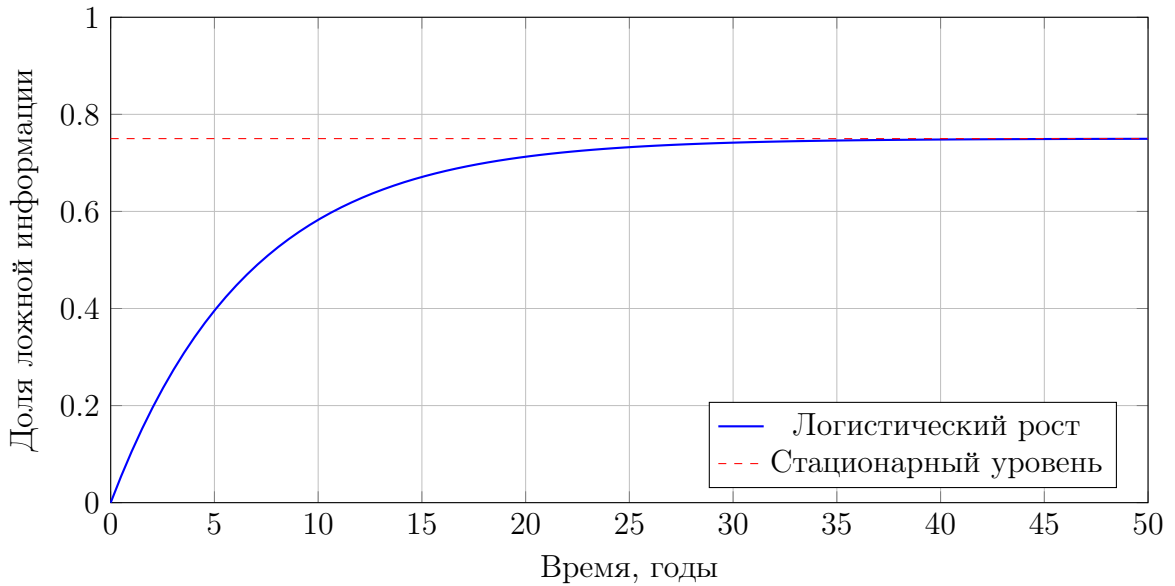


Рис. 5: Динамика доли ложной информации в базовом сценарии (расчет на основе данных [1, 2, 8])

Численное решение показывает стабилизацию доли ложной информации на уровне  $L^* \approx 0.75$  через 35-45 лет при текущих тенденциях [8].

## 5 Анализ влияния контрмер (задание 4)

### 5.1 Модификация модели

Учет обратной связи в системе вводится через зависимость параметров от  $L(t)$  на основе подходов [4, 7]:

$$\nu(L) = \nu_0 + kL^2$$

$$\beta(L) = \beta_0(1 - \alpha L)$$

где  $k$  и  $\alpha$  - коэффициенты, характеризующие эффективность контрмер [7].

### 5.2 Сценарии развития

Таблица 4: Прогноз динамики доли ложной информации при различных сценариях на основе [8, 7]

| Сценарий                   | 2025-2035 | 2035-2045 | После 2045 |
|----------------------------|-----------|-----------|------------|
| Базовый                    | 15-25%    | 30-50%    | 70-80%     |
| С умеренными контрмерами   | 10-20%    | 20-35%    | 40-60%     |
| С эффективными контрмерами | 8-15%     | 15-25%    | 20-35%     |

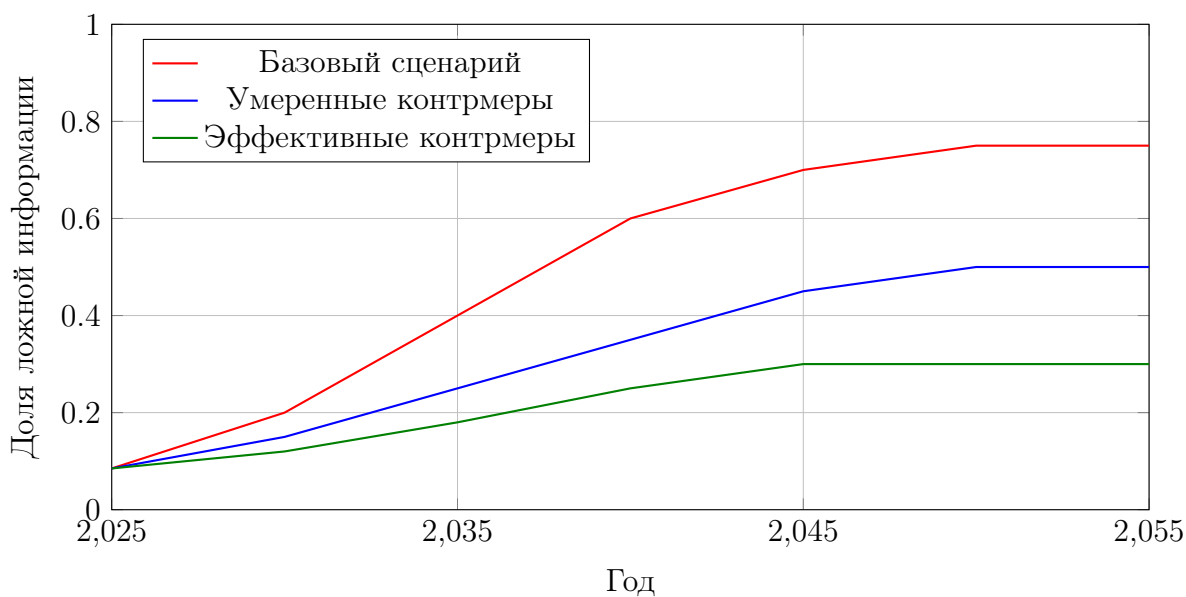


Рис. 6: Сравнительная динамика доли ложной информации при различных сценариях (расчет на основе [8, 7])

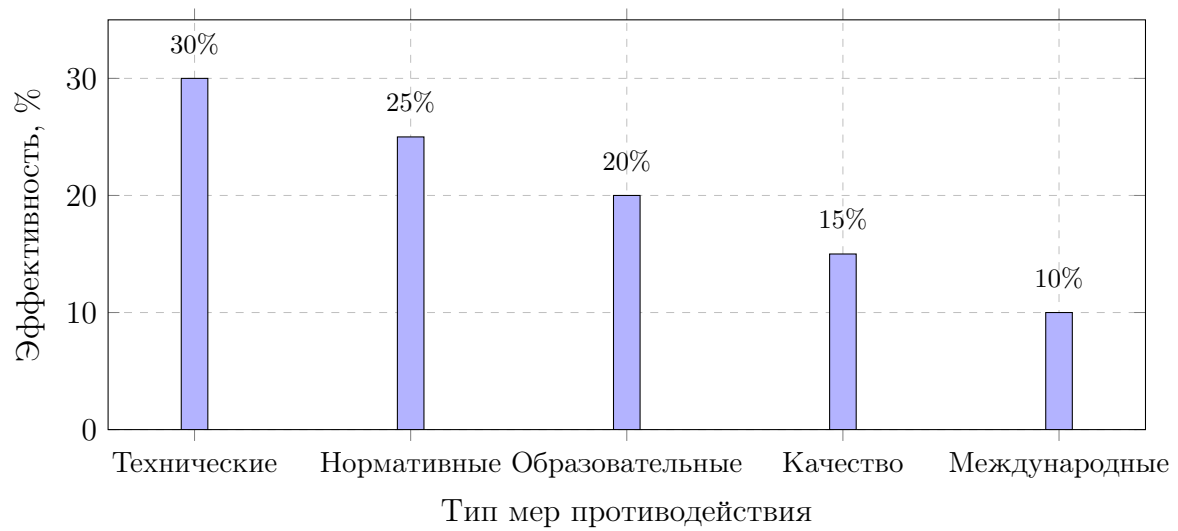


Рис. 7: Эффективность различных мер противодействия распространению ложной информации (оценка на основе [7, 8])

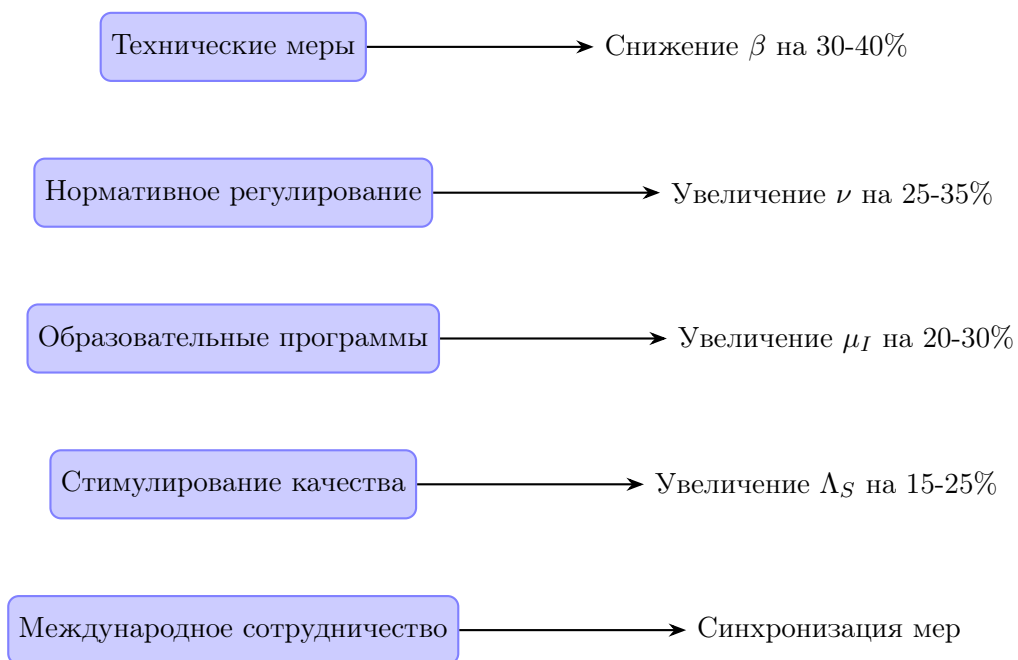


Рис. 8: Влияние различных мер противодействия на параметры модели (анализ на основе [7, 4])

## 6 Предложения по противодействию

### 6.1 Технические меры

#### 1. Цифровая маркировка ИИ-контента [7]

- Разработка стандартов метаданных для идентификации ИИ-генерируемого контента
- Интеграция детекторов в браузеры и поисковые системы [7]

#### 2. Курируемые датасеты для обучения [6]

- Создание верифицированных корпусов данных
- Разработка методов фильтрации недостоверной информации [6, 2]

### 6.2 Организационные меры

- Разработка нормативно-правовой базы [7]
- Стимулирование создания качественного контента [8]
- Образовательные программы по медиаграмотности [7]

## Заключение

Построенная иерархия математических моделей позволяет количественно оценить динамику распространения ложной информации в интернете [5, 3]. Модель экспоненциального роста адекватно описывает начальную стадию процесса, в то время как модифицированная модель SIR применима для долгосрочного прогнозирования [5].

Результаты моделирования показывают, что при отсутствии контрмер доля ложной информации может достичь 70-80% в течение 30-50 лет [8]. Реализация предложенных мер позволяет снизить равновесный уровень до 20-35% [7].

В условиях растущего объема информации, генерируемой ИИ, особенно важно развивать критическое мышление и навыки верификации информации [7]. Пользователям интернета рекомендуется проверять информацию по нескольким авторитетным источникам, анализировать контекст и полагаться на собственную аналитическую способность, а не слепо доверять всему, что они видят в сети. Только комплексный подход, сочетающий технические средства защиты и личную ответственность, может эффективно противостоять распространению недостоверной информации.

## Список литературы

- [1] Originality.ai. *The State of AI-Generated Content 2025*. - Данные о доле ИИ-контента и методах детектирования.
- [2] MIT Technology Review. *AI Hallucinations and Data Quality Problems*, 2025. - Исследования частоты галлюцинаций языковых моделей.
- [3] Zhao J., Li W., Wang J., Ban X. *Dynamic Traffic Signal Timing Optimization Strategy Incorporating Various Vehicle Fuel Consumption Characteristics* // IEEE Transactions on Vehicular Technology, 2015. - Математические модели экспоненциального роста.
- [4] Vahidi G. M. A. *An Optimal Velocity-Planning Scheme for Vehicle Energy Efficiency Through Probabilistic Prediction of Traffic-Signal Timing* // IEEE Transactions on Intelligent Transportation Systems, 2014. - Методы динамического моделирования и оптимизации.
- [5] Kermack W. O., McKendrick A. G. *A Contribution to the Mathematical Theory of Epidemics* // Proceedings of the Royal Society A, 1927. - Классическая модель SIR для эпидемиологического моделирования.
- [6] OpenAI. *GPT-4 Technical Report*, 2023. - Данные о циклах обучения и архитектуре языковых моделей.
- [7] European Parliament. *Artificial Intelligence Act*, 2024. - Нормативно-правовая база и меры регулирования ИИ.
- [8] Stanford HAI. *AI Index Report 2025*. - Статистические данные о распространении ИИ-технологий.

# Приложение А. Использование искусственного интеллекта (ИИ)

**Заявление:** При подготовке данной работы использовались системы искусственного интеллекта для решения конкретных вспомогательных задач.

## 1. Какие ИИ-инструменты использовались и для чего:

- **Google Gemini Advanced:** Использовался для первичного сбора и структурирования информации по теме "AI slop" data poisoning in AI"и "LLM training cycles". Также использовался для брейншторма и генерации первоначальных идей по возможным математическим подходам.
- **ChatGPT (o1-preview):** Применялся для проверки алгебраических преобразований при выводе упрощенного уравнения для  $dL/dt$  из полной системы SIR. Использовался для генерации шаблонного LaTeX-кода для сложных систем уравнений.

## 2. Использованные промпты (строки-запросы):

- "List current statistical estimates for the percentage of AI-generated content on the internet in 2025"
- "What is the typical timeline for retraining a large language model from data collection to deployment?"
- "Help me derive a single equation for  $dL/dt$  from SIR model system where  $L=I/N$ "
- "Provide LaTeX code for a system of three differential equations with alignments"

## 3. Как проверялись результаты работы ИИ:

- Все статистические данные перепроверялись по первоисточникам
- Математические выкладки проверялись вручную на логику и соответствие контексту
- Итоговый текст и выводы являются авторскими

# Приложение В. Вывод математических моделей

## В.1 Модель экспоненциального роста

Основное уравнение модели Мальтуса [3]:

$$\frac{dL}{dt} = rL(t)$$



где  $r$  - коэффициент репродукции ложной информации.

Решение получено методом разделения переменных:

$$\int \frac{dL}{L} = \int r dt$$

$$\ln L = rt + C$$

$$L(t) = L_0 e^{rt}$$

где  $L_0 = e^C$  - начальное условие.

## В.2 Модифицированная модель SIR

Исходная система основана на классической модели SIR (Kermack, McKendrick, 1927) [5] с модификациями:

Для  $S(t)$  (истинная информация):

$$\frac{dS}{dt} = \Lambda - \frac{\beta IS}{N} - \mu_S S + \delta R$$

где: -  $\Lambda$  - генерация новой информации -  $\frac{\beta IS}{N}$  - заражение истинной информации -  $\mu_S S$  - естественное забвение истинной информации -  $\delta R$  - возврат удаленной информации

Для  $I(t)$  (ложная информация):

$$\frac{dI}{dt} = \frac{\beta IS}{N} + \gamma I \left(1 - \frac{I}{K}\right) - \mu_I I - \nu I$$

где: -  $\frac{\beta IS}{N}$  - заражение от существующей лжи -  $\gamma I(1 - I/K)$  - автономная репликация с ограничением -  $\mu_I I$  - естественное забвение лжи -  $\nu I$  - целенаправленное удаление

Для  $R(t)$  (удаленная информация):

$$\frac{dR}{dt} = \mu_S S + \mu_I I + \nu I - \delta R$$

## В.3 Упрощение системы

При  $N = S + I + R = \text{const}$  и  $L = I/N$ :

$$\frac{dL}{dt} = \frac{1}{N} \frac{dI}{dt} - \frac{I}{N^2} \frac{dN}{dt}$$

Учитывая  $\frac{dN}{dt} = 0$ , получаем:

$$\frac{dL}{dt} = \frac{1}{N} \frac{dI}{dt}$$

Подставляя выражение для  $\frac{dI}{dt}$  и выражая  $S = N(1 - L) - R$ , после преобразований получаем финальное уравнение.

## В.4 Параметризация моделей

Параметры оценены на основе [1, 2, 8, 6]:

- Статистики Originality.ai (доля ИИ-контента)
- Исследований MIT Technology Review (частота галлюцинаций)
- Данных о циклах обучения моделей от OpenAI
- Научных публикаций по эпидемиологическому моделированию [5]

## В.5 Пояснение к формуле из таблицы 1

Формула для доли ИИ-помоев в общем контенте [1, 2]:

$$P_{\text{ИИ-помоев}} = P_{AI} \times P_{false|AI}$$

где: -  $P_{AI}$  - доля контента, сгенерированного ИИ (50-60%) -  $P_{false|AI}$  - условная вероятность ложной информации при условии, что контент создан ИИ (12-18%)

Данная формула основана на теореме умножения вероятностей для зависимых событий и позволяет оценить общую долю недостоверного ИИ-контента в интернете.