

УДК 004.852

**СИСТЕМА АДАПТИВНОГО КОНТРОЛЯ КАЧЕСТВА ОБУЧЕНИЯ МОДЕЛЕЙ ДЛЯ
КЛАССИФИКАЦИИ УЗЛОВ ЩИТОВИДНОЙ ЖЕЛЕЗЫ НА ОСНОВЕ
МОНИТОРИНГА SURPLEXITY И ОПТИМИЗАЦИИ СООТНОШЕНИЯ
КЛИНИЧЕСКИХ И СИНТЕТИЧЕСКИХ ДАННЫХ**

**Боброва Е.В. (НИЯУ МИФИ), Петровская Я.В. (НИЯУ МИФИ), Зайцев К.С. (НИЯУ
МИФИ)**

**Научный руководитель – доктор технических наук, профессор Зайцев К.С.
(НИЯУ МИФИ)**

Введение. Применение больших языковых моделей в медицинской диагностике, включая классификацию узлов щитовидной железы по системе Bethesda и EU-TIRADS, сталкивается с проблемой коллапса моделей при обучении на смеси реальных клинических данных и синтетически сгенерированных записей. Рекурсивная генерация медицинских текстов приводит к накоплению ошибок, потере разнообразия клинических паттернов и деградации точности диагностических предсказаний. Исследования показывают, что мультимодальные GPT-модели достигают точности классификации Bethesda-категорий на уровне 0.75 для Top-1 предсказаний, однако их качество зависит от баланса реальных и синтетических данных в обучающей выборке [3]. Основная проблема заключается в отсутствии механизмов динамической корректировки пропорции синтетических ультразвуковых изображений и текстовых описаний в процессе дообучения моделей на специализированных тиреоидологических корпусах. Метрика perplexity используется для оценки качества адаптации языковых моделей к медицинскому домену, при этом снижение значения демонстрирует успешную специализацию на клинической терминологии. Концепция золотого сечения как оптимального веса реальных данных (≈ 0.618) может быть адаптирована для задач классификации узлов щитовидной железы с учётом специфики EU-TIRADS [1, 4].

Основная часть. Работа направлена на разработку адаптивной системы обучения медицинских LLM с динамической коррекцией соотношения реальных цитологических заключений и синтетических данных на основе мониторинга surplexity и метрик разнообразия. Система включает специализированный LLMsurplexityMonitor, вычисляющий интегральную метрику качества как взвешенную комбинацию: (1) perplexity модели на валидационном наборе описаний узлов щитовидной железы по Bethesda-классификации; (2) diversity через distinct-1/2 метрики для оценки разнообразия терминологии ультразвуковых признаков EU-TIRADS; (3) surprise на основе KL-дивергенции распределений токенов относительно эталонного набора реальных цитологических заключений. Детектор риска коллапса анализирует тренд, волатильность и абсолютные значения surplexity в скользящем окне из пяти итераций обучения [1]. AdaptiveController корректирует mixing ratio от начального значения золотого сечения ($1/\varphi \approx 0.618$) в диапазоне [0.3, 0.9]: при превышении порога риска $collapse_score > 0.5$ доля реальных клинических данных увеличивается на $\alpha = 0.1$, что предотвращает деградацию диагностической точности при классификации Bethesda IV категории. Метод применяется для задач дифференциальной диагностики узлов по EU-TIRADS, где AI-модели достигают чувствительности 89–91% и специфичности 90–92% на реальных данных. Система интегрирована в pipeline дообучения с периодической регенерацией синтетического датасета ультразвуковых описаний и пересчётом метрик после каждой эпохи.

Выводы. Предложенный метод адаптивного управления пропорцией синтетических

медицинских данных позволяет балансировать улучшение качества классификации узлов щитовидной железы по системам Bethesda и EU-TIRADS и предотвращение коллапса модели при работе с ограниченными клиническими датасетами. Ожидается снижение collapse_score с >0.5 до <0.2 при стабилизации surplexity ≤ 5.0 после 10–15 итераций дообучения на тиреоидологических корпусах. Поддержание distinct-1 и distinct-2 метрик на уровне >0.4 и >0.3 обеспечивает сохранение разнообразия клинической терминологии в генерациях модели. Применение порога для веса синтетических данных согласуется с результатами исследований оптимальных соотношений для повышения точности классификации Bethesda-категорий с 0.75 до 0.83–0.94 для Top-3/Top-5 предсказаний. Целевые показатели включают улучшение AUC модели с 0.80 до 0.92 на валидационном наборе EU-TIRADS изображений при сохранении similarity index >0.89 для синтетических данных. Достижимость целей подтверждается экспериментами на датасетах ультразвуковых изображений узлов щитовидной железы с контролем качества через радиологическую оценку. Дальнейшее развитие включает интеграцию мультимодальных данных (US-изображения + текстовые описания) и применение в системах поддержки принятия решений для отбора пациентов на тонкоигольную биопсию.

Список использованных источников:

1. He H., Xu S., Cheng G. Golden Ratio Weighting Prevents Model Collapse // arXiv. – 2025. – arXiv:2502.18049v4.
2. Gambetta D., Gezici G., Giannotti F., Pedreschi D., Knott A., Pappalardo L. Surplexity for Mitigating Model Collapse in Generative AI // arXiv. – 2024. – arXiv:2410.12341v3.
3. Yao J., Zhou W., et al. Multimodal GPT model for assisting thyroid nodule classification with ultrasound images and clinical data // Nature Digital Medicine. – 2025. – Vol. 8. – Art. 32
4. Yao J., Chen C., et al. AI diagnosis of Bethesda category IV thyroid nodules // iScience. – 2023. – Vol. 26. – Art. 108078.