

ЭТИЧЕСКИЕ АСПЕКТЫ ПРИНЯТИЯ РЕШЕНИЙ В АВТОНОМНЫХ MAS

Автономные многоагентные системы (MAS) становятся ключевым элементом современных технологий — от умных городов до медицинской диагностики. Однако их внедрение сталкивается с этическими дилеммами, возникающими при противоречии между эффективностью (максимизацией общей полезности) и справедливостью (равномерным или приоритетным распределением ресурсов). В исследовании предложены гибридные модели, интегрирующие утилитаристские и эгалитарные подходы с использованием адаптивных весов и динамических приоритетов.

ВВЕДЕНИЕ

Развитие искусственного интеллекта и MAS приводит к активному применению автономных агентов в критически важных сферах: логистике, здравоохранении, финансах, управлении городским трафиком. Алгоритмы таких систем зачастую оптимизируют технические параметры (скорость, точность), игнорируя многокритериальность этических решений. Вопрос ответственности за несправедливость (разработчик, оператор или сама система) остаётся открытым и требует внедрения механизмов этического аудита и объяснимости (XAI).

МЕТОДОЛОГИЯ

Исследование базируется на симуляциях в условиях ограниченных ресурсов с использованием Python и платформы Mesa. Рассмотрены кейс-стади: управление трафиком беспилотников, распределение бюджетов в социальных программах и медицинское распределение донорских органов. Количественная оценка проводится с помощью таких метрик, как индекс Джини, индекс Джайнса и время отклика.

I. ТЕОРЕТИЧЕСКИЙ АНАЛИЗ ЭТИЧЕСКИХ ДИЛЕММ

Обзор литературы выявил, что утилитарный подход, направленный на максимизацию общей пользы, может приводить к маргинализации меньшинств (например, алгоритмы кредитного скоринга). Роулсианская «справедливость как честность» требует установления минимальных стандартов для всех, что зачастую снижает эффективность системы. Таким образом, необходимы адаптивные модели, способные менять приоритеты в зависимости от ситуации.

II. КРИТИКА СУЩЕСТВУЮЩИХ МОДЕЛЕЙ

Фиксированные квоты и жесткие правила нарушают гибкость MAS, а обучение на исторических

данных часто воспроизводит прошлые предубеждения. Эксперимент «The Trolley Problem» или «дилеммы трамвая» в MAS, проведенный на базе децентрализованных решений в управлении беспилотниками, показал, что децентрализация увеличивает справедливость, но замедляет реакцию системы.

III. ГИБРИДНЫЕ МОДЕЛИ БАЛАНСА

Предложены модели с динамическим перераспределением значимости критериев (адаптивные веса) и многоуровневая архитектура: на локальном уровне происходит оптимизация эффективности, а на глобальном — аудит на соответствие этическим нормам [1]. Экспериментальные данные демонстрируют, что в сценарии распределения донорских органов гибридная модель значительно снизила дискриминацию пациентов с редкими группами крови при минимальной потере эффективности. В управлении городским трафиком внедрение динамических приоритетов снизило задержки на окраинах почти наполовину без ухудшения ситуации в центральных зонах.

IV. ВЫВОДЫ

Разработка этических алгоритмов для автономных MAS требует интеграции формальных моделей распределения с адаптивными механизмами, обеспечивающими баланс между эффективностью и справедливостью. Применение гибридных моделей, основанных на количественных и качественных метриках, позволяет минимизировать социальные риски, повысить прозрачность и устойчивость систем.

1. Jiang J., Lu Z. Learning fairness in multi-agent systems // Proceedings of the 33rd International Conference on Neural Information Processing Systems. NeurIPS 2019. December 2019. P. 13854–13865. Article No.: 1241.

Кадлубай Вера Александровна, студентка ФИТиУ БГУИР, tylyla2004@gmail.com.

Научный руководитель: Хаджинова Наталья Владимировна, старший преподаватель кафедры ИТАС БГУИР, khajynova@bsuir.by.