# Краткий отчет за 2-й этап по проекту РФФИ 19-29-06099 мк

## «Разработка методов поиска уязвимостей интерфейсов взаимодействия человека с искусственным интеллектом транспортной среды «умного города»

**Руководитель проекта**
**к.т.н. Чечулин А.А.**

Современная транспортная среда "Умного города" обширна, она представляет собой совокупность общественного транспорта (поезда, метро, автобусы и троллейбусы), частного транспорта (частные машины и такси), рабочей техники (техника для уборки дорог, вывоза мусора, перевозки груза) и инфраструктуры для их навигации и функционирования (дороги, стоянки, заправочные станции и станции подзарядки, станции техобслуживания), а также людей, состоящих как из пользователей транспортной среды, так и персонала, обеспечивающего ее бесперебойную работу.

Особенностью "Умных городов" является использование искусственного интеллекта для управления транспортной инфраструктурой и оптимизации человеческих потоков. Однако, наряду с повышением эффективности и удобства управления это создает и новые угрозы для общества, так как нарушение работоспособности подобной системы управления может привести к значительному ущербу. Так, существует много различных видов атак на городские "умные" системы. Наиболее известные примеры - это компьютерные атаки на различные общественные платформы: экраны, билборды, банкоматы и другие аппараты, подключенные к интернету. К менее распространенным, но более опасным, относятся атаки на беспилотные транспортные объекты, такие как, например, поезда метро.

Данный проект направлен на исследования в области интеллектуальных транспортных систем в умном городе и решении фундаментальной задачи поиска уязвимостей интерфейсов взаимодействия человека с искусственным интеллектом. Задача данного проекта – описать и классифицировать такие проблемы безопасности (уязвимости), а также предложить пути их устранения или нейтрализации.

Результаты данного проекта помогут создателям "Умных городов" еще на стадии разработки программного и аппаратного обеспечения учесть и устранить возможные уязвимости, а если уязвимость будет обнаружена уже на этапе эксплуатации, то помогут разработать средства защиты, позволяющие ее нейтрализовать. Таким образом, исследование направлено на обеспечение безопасности организаций и граждан, живущих в "Умных городах".

В рамках второго отчетного периода были разработаны концептуальные модели интерфейсов взаимодействия оператор-система, основанная на технологиях машинного зрения и визуализации данных и система-оператор, основанная на технологиях визуализации, учитывающих когнитивный аппарат человека. Кроме того, были разработаны методы определения типа интерфейса человек - искусственный интеллект и методы поиска уязвимостей в интерфейсах взаимодействия пользователь - система и система-пользователь. Полученные теоретические результаты были частично реализованы в рамках компонентов программного обеспечения для программно-аппаратного стенда с целью последующего проведения экспериментов.

1) Концептуальная модель интерфейса взаимодействия оператор-система

Была разработана концептуальная модель интерфейса взаимодействия оператора с системой, которая описывает основные процессы, которые поддерживает такой интерфейс взаимодействия. Представленные в рамках концептуальной модели процессы позволяют оператору автоматизировать мониторинг водителей. Для отслеживания состояния усталости оператора был предложен процесс «Обнаружения усталости оператора», который отслеживает уровень его усталости при помощи RGB камеры и предупреждает его об этом для предотвращения опасных ситуаций.

2) Концептуальная модель интерфейса взаимодействия система-оператор

Была разработана концептуальная модель интерфейса взаимодействия система-оператор, основанной на технологиях визуализации, учитывающих когнитивный аппарат человека, которая состоит из следующих промежуточных результатов:

- Модель когнитивной нагрузки оператора. Была разработана модель когнитивной нагрузки, которая состоит из двух параметров: точности интерпретации данных и скорости принятия решений. В зависимости от требований к анализу данных, задается приоритетное соотношение скорости и точности. В зависимости от заданного приоритета, в интерфейсе могут быть задействованы различные модели визуализации и методы взаимодействия оператора с системой, обеспечивая оптимальный уровень когнитивной нагрузки.

- Интерфейсы взаимодействия, которые ранжированы по параметрам когнитивной нагрузки. Исходя из параметров когнитивной нагрузки данных моделей, будет приниматься решение об целесообразности использования того или иного интерфейса взаимодействия. Всего были рассмотрены следующие интерфейсы взаимодействия и их параметры: нажатие на объект или область, печать на клавиатуре, нажатие и проведение пальцем в другую часть экрана для выделения области или увеличения области, нажатие на объект и перетаскивание его в другую область экрана, перемещения указателя из одного угла экрана в другой, сведение/разведение двух или более пальцев для увеличения объекта на экране, поворот объекта вокруг своей оси двумя пальцами на экране, следование какой-либо траектории, листание снизу вверх или сверху вниз.

- Модели визуализации данных оператором, которые ранжированы по параметрам когнитивной нагрузки. Исходя из параметров когнитивной нагрузки данных моделей, оценить целесообразность использования той или иной модели визуализации. Всего были рассмотрены следующие модели визуализации и их параметры: тепловые карты, облака слов, простейшие графики (линейные, столбчатые, круговое и т.д.), графики рассеивания, матрицы и модели для многомерных структур, параллельные координаты (и им подобные радиальные координаты), карты деревьев (включают упаковки шаров, карты деревьев вороного), графы (деревья, радиальные деревья) и модели для планарных структур, Карты

вороного и модели для неструктурированных структур, Диаграммы хорда (радиальный граф и радиальное дерево), а также многие совмещенные модели, Карты (стран, городов, помещений и т.д.), Диаграммы Вороного.

3) Метод определения типа интерфейса человек – искусственный интеллект

Был разработан метод определения типа интерфейса человек – искусственный интеллект, базирующийся на анализе предметной области и работающий в соответствии с моделью, разработанной в течение первого года выполнения проекта. Данный метод представляет собой последовательный переход от класса компонента к классу его взаимодействия, а от класса взаимодействия к классу интерфейса на основе его смысловой ориентированности.

Идея полученного результата заключается в построении метода определения типа интерфейса человек – искусственный интеллект на основе требований к “необходимости и достаточности”, что позволяет гарантировано отнести даже гипотетический интерфейс к одному из классов. Также представленный метод работает с разработанной в течение первого года выполнения проекта моделью, в которую заложена возможность ее расширения за счет добавления новых пар категориального деления. Подобное расширение позволяет повышать уровень детализации классификации.

4) Метод поиска уязвимостей в интерфейсах взаимодействия пользователь

Был разработан метод обнаружения уязвимостей в рамках взаимодействия человек – система в рамках системы мониторинга водителя, который объединяет две основные части: последовательность обнаружения уязвимостей и вспомогательный цикл.

Для создания моделей машинного обучения требуются наборы данных, которые содержат информацию в той же форме, что и на этапе сбора данных. Набор данных обычно делится на три подмножества: обучающий набор (используемый для обучения моделей машинного обучения и определения пороговых значений), валидационный набор (используемый для проверки моделей машинного обучения и пороговых значений) и тестовый набор (используемый для оценки качества обученных моделей и пороговых значений. И валидационный, и тестовый наборы используются для оценки качества модели. Разница заключается в следующем: валидационный набор используется для эмпирического выбора структуры модели, архитектуры нейронной сети, гиперпараметров обучения и т. д. Другими словами, он может использоваться для полного описания решения задачи, основанного на методах обучения. Тестовый набор используется только для оценки качества законченной модели на данных, недоступных в процессе ее разработки, выбора и обучения.

5) Метод поиска уязвимостей в интерфейсах взаимодействия система-пользователь.

Метод позволяет определять уязвимости, присущие интерфейсам по их признакам. Метод базируется на разработанной модели классификации уязвимостей и также имеет аналитический вид.

Принцип работы метода основан на преобразовании признаков исследуемого интерфейса в класс интерфейса, затем во множество модулей интерфейсов, затем во множество уязвимостей элементов интерфейса и затем во итоговое множество уязвимостей интерфейса.

6) Компоненты программного обеспечения для программно-аппаратного стенда

Ряд программных прототипов были разработаны и зарегистрированы в Реестре программ для ЭВМ в рамках выполнений проекта.

В рамках второго года выполнения проекта была опубликована новость в издании Научная Россия от 22.05.2021 «Ученые СПб ФИЦ РАН разработали систему защиты для транспортной инфраструктуры "умного города" (https://scientificrussia.ru/articles/uchenye-spb-fits-ran-razrabotali-sistemu-zashchity-dlya-transportnoj-infrastruktury-umnogo-goroda) и в ТАСС Наука от 27.01.2022 (https://nauka.tass.ru/nauka/13541465).

В рамках отчетного периода были опубликованы статьи в ведущих профильных журналах в России (Системы управления, связи и безопасности, Информатизация и связь, Системы анализа и обработки данных), и за её пределами (одна в журнале Electronics, входящем во второй квартиль системы цитирования Scopus и одна в журнале Sensors, входящим в первый квартиль системы цитирования Web of Science). Кроме того, на рассмотрении в журналах, входящих в первый квартиль системы цитирования Web of Science находятся еще 2 статьи по данному проекту. Результаты проекта также были представлены на двух конференциях в России и трех иностранных конференциях.