Ройзензон Г.В.

Формализация понятия этики в ИИ

В докладе рассматриваются основные проблемы формализации понятия этики в ИИ. Одним из возможных путей решения поставленной задачи является разработка норм (этических критериев) для каждого из проектов стандартов этики ИИ (P7000–P7013), которые были предложены в рамках глобальной инициативы IEEE по этически обусловленному проектированию в 2016–2018 гг. Проблема формализации этических норм тесно связана с более общей задачей, а именно: с формализацией гуманитарного знания и включает в себя две основные задачи. Первая — это создание форм представлений норм, вторая — выбор соответствующего математического аппарата для работы с этими формами: сопоставления, измерения, анализа и т.д. К формальным методам, использование которых весьма востребовано для поставленной в докладе задачи, можно отнести следующие: булева алгебра, многозначные логики, нечетная логика и теория вероятностей, теория решеток, а также методы вербального анализа решений (ВАР). В рамках доклада проведен критический анализ (достоинства и недостатки) каждого из рассматриваемых формальных подходов (математического инструментария). В качестве примеров рассмотрены варианты формализации понятия этики для проектов стандартов IEEE P7011 и P7013 с использованием метода ВАР ЦИКЛ. В частности, проект стандарта этики ИИ P7011 ориентирован на оценку надежности новостных источников. Последнее десятилетие характеризуется беспрецедентными масштабами ведения различных информационных войн. Проект стандарта IEEE этики ИИ P7011 (оценка надежности новостных источников) ориентирован на устранение негативных последствий неконтролируемого распространения поддельных («фейковых») новостей путем предоставления открытой системы оценок. Очевидно, что «фейковые» новости являются одним из важнейших инструментов информационных войн, поэтому любые инициативы, направленные на «оздоровление» информационного пространства представляют определенный интерес. Для решения указанной задачи, очевидно, могут быть использованы методы лингвистической семантики и семантического анализа текстов (семантические технологии Web). Тем не менее, в современных условиях при распространении «фейковых» новостей необходимо проводить анализ и других компонентов (например, видеоряд, фотографии, звук и т.п.), т.к. они также могут быть фальсифицированы. В докладе представлена система критериев для оценки надежности новостных источников, которая, в определенных условиях, может также рассматриваться и как система этических норм в рамках стандарта этики ИИ P7011. В рамках проекта стандарта IEEE P7013 исследуются различные технологии автоматизированного анализа состояния лица. В современной банковской практике клиенты все чаще сталкиваются с ситуацией, когда вместо аналитиков банка, по существу решение о выдачи кредита принимают различные компьютерные системы, использующие технологии ИИ. В ряде случаев это приводит к различного рода неясностям и даже противоречиям. Особый интерес представляют такие системы оценки кредитного риска (пример использования такой системы представлен в докладе) в связке с проектом стандарта IEEE P7013 применительно к практике выдачи крупных займов корпоративным заемщикам. В современных условиях в банковском деле все чаще применяются компьютерные системы, использующие технологии ИИ, для анализа мимики лица и поведения человека (например, представителя заемщика) на переговорах (фактически осуществляется анализ характерных повадок мошенников и т.п.). Именно для сертификации подобных компьютерных систем, осуществляющих анализ состояния лица, в которых используются различные технологии ИИ, и требуется повсеместное внедрение стандарта IEEE P7013 этики ИИ.