УДК 004.83

ЧЕЛОВЕК УПРАВЛЯЕТ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ ИЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ УПРАВЛЯЕТ ЧЕЛОВЕКОМ?

Пылов П.А., Кудаева И.В., студенты группы ИТм-201, I курс
Научный руководитель: Баумгартэн М.И., канд. физ.-мат. наук, доцент
Кузбасский государственный технический университет имени Т.Ф. Горбачева г. Кемерово

В бурно развивающемся современном мире, полном цифровизации, который уже совсем невозможно представить без сети Интернет, совсем скоро может настать момент создания полноценного «сильного» искусственного интеллекта [1].

Многочисленные группы в социуме возлагают большие надежды на искусственный интеллект, как технологию, которая позволит решить рутинные человеческие задачи, позволяя людям уделять больше времени своим повседневным и более приятным жизненным делам. Правда, не разобранным остаётся тот факт, как будут распределяться ресурсы (заработная плата и другое) при замене большинства функций людей искусственным интеллектом. Предполагается, что системы искусственного интеллекта будут работать, отдавая большую часть своей прибыли социуму [2]. Однако, даже если следовать такой оптимистичной логике, то от нас всё равно никуда не пропал вопрос безопасности таких систем.

Люди настолько привыкли к сети Интернет и всем преимуществам, которые она предоставляет, что уже не представляют свою жизнь отдельно от неё. Но немногие вспоминают о том, что большинство физических устройств уже оснащены различными автоматическими модулями и имеют доступ в Интернет. Поэтому контекст безопасности, в случае рассматривания систем искусственного интеллекта, касается не только информационной составляющей, но и физической, поскольку, получив полный доступ к сети Интернет, можно взять под контроль и все устройства, управляемые с его помощью [3].

Каким образом можно обезопасить человечество в целом от потенциально возможного нападения искусственного интеллекта на электронное сетевое пространство, защитив тем самым и физическую составляющую? Этот вопрос давно стал тривиальным, но, тем не менее, вразумительный и исчерпывающий ответ на него был сформулирован лишь к концу 2020 года на научно-инженерном форуме корпорации Майкрософт [4]. Объединённым коллективом специалистов был сформулированы основные тезисы, которые сводятся к тому, что тестирование систем «сильного» искусственного интеллекта возможно строго в локальной среде [4].

Это означает, что испытывать такие системы на устройствах, имеющих доступ в сеть Интернет категорически запрещается. Кроме этого, настоятельно рекомендуется исключать любую аппаратную возможность

доступа в Интернет [4]. Это означает, что, например, перед тем как испытывать такую систему на современном ноутбуке, рекомендуется не только отсоединить кабель сети Интернет, но и аппаратно извлечь модуль беспроводной Wi-Fi и Bluetooth связи с внешними системами [4].

Поразмыслим над тем, насколько это необходимо. Для начала, следует понимать, что такая рекомендация носит характер предупреждающей защиты, то есть обеспечивает защиту от потенциальной угрозы. К такой угрозе относится возможность неконтролируемого выхода системы искусственного интеллекта в Интернет.

Чем это может быть опасно? Люди не имеют опыта в принципах работы искусственного интеллекта, а посему не могут быть уверенными, что искусственный разум сопоставим с человеческим (он может превосходить его в разы) [5]. В таком случае, даже если отсутствует подключение к беспроводной сети, то система может самостоятельно подобрать комбинацию пароля к любой доступной сети Интернет и получить свободный доступ к ней. Такая ситуация означает, что, теоретически, алгоритм может копировать свои файлы и неконтролируемо распространяться в Интернете, осуществляя при этом любые возможные действия.

Возвращаясь к главному вопросу статьи, проясним: кто же на самом деле является ведущим субъектом, а кто – ведомым. Сохраняя тенденцию общепринятых норм и правил, можно уверенно заявить, что полное отсутствие у системы доступа к общемировой сети, а также невозможности наладить связь с другими персональными компьютерами, на корню ограничивает возможности информационной системы.

Концепция, предлагающая запуск искусственного интеллекта, на компьютере, не имеющем доступа в сеть Интернет, называется виртуальной средой (локальная сеть), на рисунке 1 представлена ключевая особенность такой реализации – самое тяжелое последствие, которое может возникнуть в данной среде есть непосредственное крушение среды локальной сети. То есть, самый тяжелый случай – это выход из строя максимум пяти, представленных на рисунке 1 (левая часть – локальная сеть), персональных компьютеров и данных, хранящихся на них. С точки зрения социума и даже отдельного человека такой урон является ничтожным по сравнению с общемировой сетью Интернет (правая часть рисунка 1 – глобальная сеть).

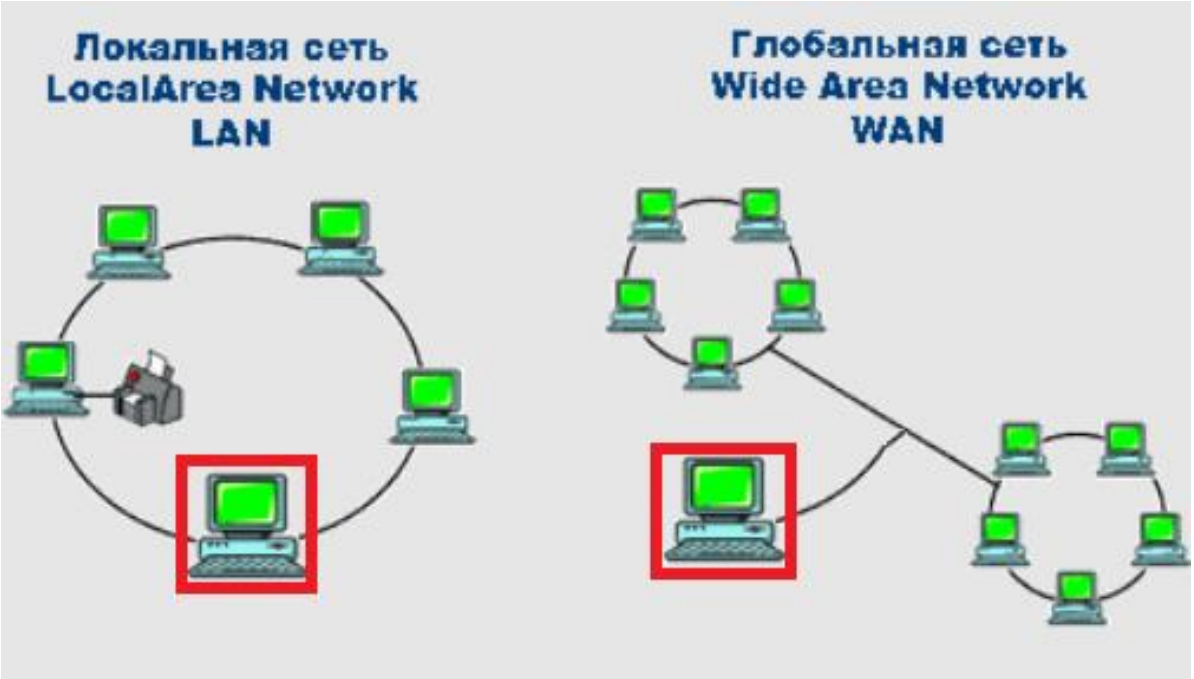

Рисунок 1 – Примеры запусков искусственного интеллекта в разных условиях

Кроме этого, поскольку человек сам определяет настройки персонального компьютера, то только он, в конечном счёте, является его непосредственным управителем. Соответственно, такой подход гарантированно оставит человека, сохраняя за ним всю прерогативу действий и желаний, ведущим субъектом над искусственным интеллектом.

Исходя из этого, следует, что, если человек хочет управлять искусственным интеллектом и предопределять его действия, а не стать ведомым субъектом, то необходимо использовать конфигурацию отдельно взятого компьютера, отключенного от сети Интернет. Но, учитывая, что ресурсов современных компьютеров недостаточно для функционирования такого сложного алгоритма, отметим, что предполагаемым вариантом для решения задачи является использование конфигурации локального сервера (очень мощного компьютера, не имеющего выход во Всемирную сеть). Примером такой конфигурации является схема, изображенная на рисунке 2.

Рисунок 2 – Конфигурация сети для контролируемой человеком работы искусственного интеллекта

Подводя итог вышесказанному, следует обозначить, что в современном мире, учитывая весь его огромный спектр возможностей в сфере информационных технологий, однако, всё ещё не хватает потенциала отдельно взятого персонального компьютера. Бесспорно, что с ростом научно-технического прогресса, уже в самом ближайшем будущем миру предстанут новые вычислительные характеристики персональных компьютеров, использование которых позволит безопасно испытать алгоритм, который по праву считается философским камнем информационных технологий – искусственный интеллект.

Список литературы:

1. Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии. / Н. Бостром. - Текст: непосредственный // М.: Манн, Иванов и Фербер. - 2020. - 496 с.
2. Харпер Р., Родден Т. Роджерс И. Селлен И. Быть человеком. Взаимодействие человека и компьютера в 2020 году – Текст: электронный // Интернет-хранилище. – URL: http://download.microsoft.com/documents/rus/devcenter/MSR_Being_Human_HCI_2020_rus.pdf (дата обращения: 12.02.2021).
3. Технологии безопасности в эпоху искусственного интеллекта [Электронный ресурс]. – Режим доступа: https://www.anti-malware.ru/analytics/Technology_Analysis/using-artificial-intelligence-technologies-in-information-security.
4. Нормативные регулирующие правила использования и испытаний систем искусственного интеллекта [Электронный ресурс]. – Режим доступа:

https://docs.microsoft.com/ru-ru/security/engineering/securing-artificial-intelligence-machine-learning

5. Петрунин, Ю. Ю. Философия искусственного интеллекта в концепциях нейронаук. / Ю. Ю. Петрунин, М. А. Рязанов, А. В. Савельев. – Текст: непосредственный // М.: МАКС Пресс. - 2010. - 187 с.