кращий стан з цієї пари, який порівнюється з іншим випадковим станом за такою же процедурою. Наприкінці залишається тільки один найкращій стан об'єкта/процесу.

Висновок. Методи вербального аналізу виявляються ефективними для синтезу поясненних моделей ШІ, що включає кілька етапів: визначення системи понять, створення критеріальних описів станів, їх класифікація, впорядкування та обрання найкращого стану. Вони підкреслюють важливість використання лінгвістичної інформації разом з числовими даними для комплексного аналізу складних проблем.

Список використаних джерел

1. Mishra P. Practical Explainable AI Using Python. Apress Berkeley, CA; 2022: 344.
2. Фастовський ЕГ, Єльчанінов ДБ. Інформаційна технологія аналізу та синтезу поясненних моделей штучного інтелекту. У: Теоретичні та практичні дослідження молодих вчених [Інтернет]; 28-30 лист. 2023; Харків. Харків: НТУ "ХПІ"; 2023 [цитовано 29 лют. 2024]. с. 85. Доступно на: https://repository.kpi.kharkov.ua/handle/KhPI-Press/71485
3. [24]7.ai [Internet]. Conversational AI: What Is It, How Does It Work, and Why Does It Matter?; [cited 2024 Feb 29]. Available from: https://www.247.ai/insights/conversational-ai-what-it-and-how-does-it-work
4. Moshkovich H, Mechitov A, Olson D. Verbal Decision Analysis. In: Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science, vol 78. Springer, New York, NY; 2005 [cited 2024 Feb 29]. p. 609–633. Available from: https://doi.org/10.1007/0-387-23081-5_15
5. Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. The Breast, Volume 50; 2020: 49–55. Available from: https://doi.org/10.1016/j.breast.2020.01.008

УДК 004.89

ВИКОРИСТАННЯ СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІДЕНТИФІКАЦІЇ РИЗИКІВ, ЩО ВИНИКАЮТЬ ЯК НАСЛІДОК ВИКОРИСТАННЯ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ.

Ширшов Р.А., науковий співробітник, signorum@gmail.com,
Національна академія Служби безпеки України

Використання системи довірених штучного інтелекту (далі – ШІ) для управління ризиками, які виникають внаслідок використання інших (цільових) систем ШІ, буде ефективним засобом забезпечення їх безпеки. Використання довірених систем ШІ може включати в себе наступне:

1. Ідентифікація ризиків. Аналіз архітектури системи та алгоритмів, що використовуються в оцінюваній системі, аналіз типів даних та інших параметрів. Система може використовувати раніше визначену інформацію про загрози, вразливості та сценарії їх реалізації та на її основі визначати

слабкі місця, потенційні точки виникнення помилок або точки, де система може взаємодіяти з зовнішнім світом.

2. Моніторинг поведінки: Система управління ризиками на базі ШІ може стежити за поведінкою та результатами інших систем ШІ, аналізуючи результати їх діяльності, та даючи оцінку відповідності заздалегідь визначеним метрикам та коректність поведінки в режимі реального часу.
3. Автоматичне виявлення аномалій: Використовуючи методи машинного навчання, система виявлятиме аномалії або відхилення від норми (значень метрик) в поведінці систем ШІ.
4. Прогнозування ризиків: На основі історичних даних можливо здійснювати прогноз щодо потенційних проблем або непередбачуваної поведінки в майбутньому.
5. Автоматична корекція/інформування: У випадках виявлення ризику, система може автоматично вносити корективи в роботу контрольованої системи ШІ, наприклад, змінюючи параметри її функціонування, обмежуючи її дії та/або надсилаючи відповідну інформацію зацікавленим в процесі контролю та керування сторонам.
6. Виконання сценаріїв "чорної скриньки": Для аналізу поведінки системи ШІ можна використовувати раніше визначені сценарії, де система ШІ тестується у контрольованому оточенні.
7. Аналіз причинно-наслідкових зв'язків: Система може допомогти аналізувати причини певної поведінки системи ШІ, визначаючи, чи була ця поведінка результатом вхідних даних, алгоритмів, або інших факторів чи, можливо, зовнішнього впливу.
8. Зворотний зв'язок і навчання: На основі аналізу ризиків та інцидентів система буде навчатися, вдосконалюючи методи виявлення та реагування на ризики.

Висновки

Виявлення загроз та оцінки ризиків, пов'язаних з розвитком і впровадженням систем штучного інтелекту, має бути розглянуте з урахуванням всіх аспектів впровадження та використання таких систем. Для забезпечення якісного управління ризиками ШІ реалізація заходів за нормативно-правовим технічним, організаційним та напрямками, як це запропоновано нижче.

Нормативно-правові заходи

1. Визначення та прийняття державної політики в галузі штучного інтелекту[1].
2. Створення чітких законодавчих норм, які регулюють розробку та використання ШІ, прийняття законів, які встановлюють стандарти безпеки для ШІ в кібернетичних системах [2].
3. Участь в міжнародних угодах та ініціативах, спрямованих на регулювання ШІ, участь у створенні міжнародних норм і стандартів безпеки для ШІ та забезпечення їх використання в Україні після відповідної оцінки та гармонізації.

Технічні заходи

1. Розробка та впровадження безпечних систем ШІ, що включають вбудовані заходи безпеки, для протистояння атакам та спробам впливу.
2. Встановлення технічних обмежень на доступ ШІ до сенситивних даних, таких як персональні дані громадян, може допомогти запобігти неправильному використанню цих даних.
3. Створення систем ШІ для проведення глибокого аудиту знань прикладних систем ШІ.
4. Розробка механізмів виявлення ознак роботи небезпечних ШІ
5. Розробка заходів з активної протидії небезпечним ШІ

Організаційні заходи

1. Створення моделі управління ризиками ШІ, яка має містити механізми визначення рівнів загроз та імовірності їх реалізації в різних областях діяльності людини, суспільства, держави.
2. Проведення освітніх кампаній для збільшення обізнаності про потенційні ризики, пов'язані з ШІ, може допомогти в запобіганні їх використанню з недоброякісними намірами.
3. Співпраця з приватним сектором для створення безпечних систем ШІ і розробки ефективних стратегій протидії потенційним загрозам.
4. Створення спеціалізованих органів, які будуть відповідальні за моніторинг та реагування на загрози, пов'язані з ШІ.
5. Обмеження доступу до масивів даних та спеціалізованих баз знань створених державними установами для використання їх в моделях навчання штучного інтелекту.

Список використаних джерел

1.Пацурія Н. Впровадження технологій штучного інтелекту у забезпечення національної безпеки та обороноздатності України: проблеми та перспективи повоєнного періоду: веб-сайт. URL: https://coordynata.com.ua/vprovadzenna-tehnologij-stucnogo-intelektu-u-zabezpecenna-nacionalnoi-bezpeki-ta-oboronozdatnosti-ukraini-problemi-ta-perspektivi-povoennogo-periodu (дата звернення: 26.02.2024).

2.Чайковська В., Губа Р. Європарламент схвалив план регулювання штучного інтелекту: веб-сайт. URL: https://www.dw.com/uk/evroparlament-shvaliv-plan-reguluvanna-stucnogo-intelektu/a-65912171 (дата звернення: 26.02.2024).