УДК 340

КЕРІВНИЦТВО З ЕТИКИ ДЛЯ ШІ, ЩО ЗАСЛУГОВУЄ ДОВІРИ (ETHICS GUIDELINES FOR A TRUSTWORTHY AI)

Клименко Д.А.
Науковий керівник – канд. юрид. наук, доц. Турута О.В.
Харківський національний університет радіоелектроніки, каф. філософії,
м. Харків, Україна
тел. +38(098) 986-05-44

The world is constantly changing, and society is on the verge of significant changes related to scientific and technological development. Artificial intelligence and robots are no longer just the stuff of science fiction but now are commonplace. Are people ready for harmonious coexistence with AI? Do we have a coherent ethical and regulatory system for controlling smart machines? This text highlights the importance of not only the technical development of AI, but also the importance of regulatory frameworks to ensure that AI systems are safely introduced into all areas of life, minimizing the negative consequences, and maximizing the benefits of AI.

Існування штучного інтелекту (надалі – ШІ) більше не здається предметом з області фантастики, а є цілковито реальним явищем в нашому житті завдяки покращенню комп'ютерного устаткування та поглибленню знань у сфері нейронних мереж. Як стверджують вчені, використання технологій ШІ здатне покращити людське існування у всіх його аспектах [1, с. 16]. Проте, постає логічне питання, чи не можуть розумні машини створити смертельної загрози для всього світу?

Дійсно, історія людства знає чимало випадків, коли вчені, що були одержимі своїми працями, не замислювались над наслідками та впливом на навколишнє середовище чи суспільство своїх винаходів, вважаючи власні дії суто технічними. Насправді важливо усвідомлювати соціальну та моральну відповідальність у своїй роботі, до чого й закликає етика.

Відносно встановлення етичних норм для функціонування систем штучного інтелекту, треба констатувати, що цей процес є досить важким і тривалим, саме через складність структури людських цінностей, їх абстрактність та неоднозначність. Для прикладу розглянемо країни Європи. Як і більшість високорозвинених країн вони мають наміри впровадити ШІ в приватний і державний сектор, аби збільшити економічну продуктивність, що зберігало б сталий розвиток, але одночасно існує недовіра до цих інновацій через низку нещасних випадків, пов'язаних з технологіями ШІ, що трапилися за останні декілька років. Тож, очевидно, що будь-який винахід має підлягати нормативному регулюванню для створення безпечних умов й мінімізації негативних наслідків користування відповідними технологіями, тому у червні 2018 року Європейська комісія зібрала експертну групу високого рівня з ШІ (AI HLEG), яка опублікувала

Керівництво з етики для ШІ, що заслуговує довіри (Ethic Guidelines for Trustworthy AI) у квітні 2019. Цей документ затверджує основні принципи «надійного ШІ» та можливі методи реалізації всіх положень [2]. Основними ознаками системи ШІ, що заслуговує довіри є її орієнтованість на людей (human-centric AI), а також наявність трьох компонентів, яких слід дотримуватися протягом усього життєвого циклу системи:

1. Законність, тобто відповідність всім чинним законам і правилам;

2. Етичність, забезпечення дотримання етичних принципів і цінностей;

3. Надійність як з технічної, так і з соціальної точки зору, оскільки, навіть маючи добрі наміри, системи штучного інтелекту можуть завдати ненавмисної шкоди.

В документі експертна група окремо виділяє чотири етичних принципи, як повага до прав і свобод людини, запобігання насиллю, справедливість та зрозумілість алгоритмів «надійного ШІ», що засновані на фундаментальних правах поваги до демократії, закону та людини.

Для формування довіри до ШІ існують технічні методи, що полягають у подальшому розвитку інженерних систем, що мають обов'язково спиратися на нормативно-правову базу, яка є складником нетехнічних методів реалізації «надійного ШІ». Усі етичні принципи, що були зазначені вище, транслюються в конкретні вимоги для досягнення «безпечного ШІ». Вони включають такі системні, індивідуальні та суспільні аспекти, як людський нагляд, відстеження алгоритмів, прозорість коду, технічна надійність та безпека, стійкість до атак, наявність запасного плану, конфіденційність, повага до особистих даних, уникнення несправедливої упередженості, соціальний та екологічний добробут, тобто гармонійне впровадження технологій у суспільні рамки моралі [2].

Отже, публікація Керівництва з етики для ШІ, що заслуговує довіри, хоч не і не остаточний, але однозначно важливий крок для країн Європи на шляху врегулювання штучного інтелекту для запобігання якомога більшої кількості негативних наслідків при отриманні максимальних переваг використання новітніх технологій. У будь-якому випадку, як наголошує професор Оксфорду Нік Бостром: тільки від нас залежить, до кращого чи гіршого призведе поява розумних машин [1].

Список використаних джерел

1. Бостром Н. (2020). Суперінтелект: стратегії і небезпеки розвитку розумних машин: пер. с англ. / Нік Бостром; пер. Антон Ящук, Антоніна Ящук; кер. проекту Галина Харук-Бачуро; наук. ред. Тетяна Манжос. – Київ: Наш Формат.

2. Ethics guidelines for a trustworthy AI. Режим доступу: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai