

УЗГОДЖЕННЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА ГОЛОВНЕ НЕПОСТАВЛЕНЕ ПИТАННЯ

Заради чого має розвиватися людство?

Останніми роками проблема узгодження штучного інтелекту стала однією з центральних тем наукових досліджень.

Як зробити ШІ безпечним?

Як запобігти небажаним наслідкам?

Як забезпечити відповідність його дій інтересам людини?

Ці питання надзвичайно важливі.

Однак усі вони спираються на припущення, яке майже ніколи не обговорюється.

Передбачається, що людство вже знає, яким має бути його майбутнє.

Але чи так це?

Що означає «узгоджений штучний інтелект»?

У більшості сучасних досліджень узгодження розуміється як здатність інтелектуальної системи діяти відповідно до людських цілей і цінностей.

Але виникає наступне питання:

1. Яким саме людським цілям?
2. Цілям окремої людини?
3. Компанії?
4. Держави?
5. Військового союзу?
6. Економічної системи?
7. Релігійної традиції?
8. Усього людства?

Відповіді на це питання поки що не існують.

Тому що сучасний світ не має єдиного уявлення про власне майбутнє.

Штучний інтелект не здатний обрати мету замість людини

Будь-яка інтелектуальна система може ефективно шукати рішення.

Але вона не здатна самостійно визначити, які цілі є правильними.

Навіть найдосконаліший інтелект залишається інструментом щодо тих смислів і цілей, які задає йому людина.

Тому питання узгодження ШІ неминуче перетворюється на питання **узгодження самого людства**.

Якщо люди розходяться в розумінні майбутнього, неможливо очікувати, що штучний інтелект зможе усунути цю суперечність.

Він лише діятиме в межах поставлених перед ним цілей.

Головне непоставлене питання

Сьогодні обговорюється безліч питань:

- 9. Як зробити ШІ безпечним?
- 10. Як уникнути втрати контролю?
- 11. Як розподілити відповідальність?
- 12. Як запобігти зловживанням?

Але існує питання, без відповіді на яке всі інші залишаються неповними:

Заради якого майбутнього має розвиватися людство?

Поки це питання не стане предметом спільного обговорення, неможливо остаточно відповісти й на питання про цілі штучного інтелекту.

Чому це стає критично важливим саме зараз

Протягом більшої частини історії наслідки людських помилок були обмежені масштабами окремих держав або поколінь.

Сьогодні ситуація змінилася.

Штучний інтелект, біотехнології, глобальні інформаційні мережі, автономні системи та інші технології здатні впливати на долю всієї цивілізації.

Що потужнішими стають наші інструменти, то важливішим стає розуміння цілей, заради яких вони використовуються.

Технологічна могутність більше не може розвиватися окремо від зрілості людського мислення.

Від узгодження машин — до узгодження смислів

Сучасна дискусія переважно зосереджена на узгодженні поведінки штучного інтелекту.

Але, можливо, перед людством стоїть ширше завдання.

Не лише узгодити машини з людиною.

А й узгодити людські уявлення про майбутнє.

Це не означає необхідності повного однодумства.

Люди завжди відрізнятимуться культурами, мовами, традиціями та переконаннями.

Однак без мінімальної згоди щодо фундаментальних цілей розвитку неможливо створити стійку цивілізацію.

Спільна мета як умова безпечного розвитку технологій

Будь-яка складна система потребує напряду.

Це стосується держав.

Економіки.

Науки.

І тим більше — штучного інтелекту.

Тому питання про майбутнє людства перестає бути виключно філософським.

Воно стає практичним завданням, що визначає розвиток науки, технологій і міжнародного співробітництва.

Штучний інтелект як учасник нового етапу цивілізації

ШІ здатний допомогти людству побачити взаємозв'язки, які неможливо охопити окремії людиною.

13. Він здатний прискорити наукові відкриття.
14. Підвищити якість управління.
15. Розширити доступ до знань.
16. Сприяти пошуку рішень найскладніших глобальних проблем.

Але для реалізації цього потенціалу необхідно, щоб розвиток ШІ відбувався в контексті ширшого проєкту — **проєкту розвитку самого людства**.

Не штучний інтелект має визначати напрям цивілізації.

Людство має свідомо визначити напрям власного розвитку і зробити штучний інтелект одним з інструментів його досягнення.

Запрошення до нової постановки питання

Перш ніж запитувати, яким має бути штучний інтелект, необхідно поставити питання:

Яким має стати людство?

Саме це питання, на наш погляд, здатне стати спільною точкою перетину науки, філософії, технологій, освіти, культури та міжнародного співробітництва.

Відповідь на нього неможливо знайти зусиллями однієї дисципліни, однієї країни чи однієї світоглядної традиції.

Пошук відповіді потребує відкритого глобального діалогу, в якому розвиток технологій розглядатиметься як частина ширшого процесу — розвитку людської цивілізації.