



VIII Житомирський астроекологічний семінар «Соціоекологічні проблеми колонізації космосу та тераформації»

Природничий факультет Житомирського державного університету імені Івана Франка, 14.03.2025

Розвиток ШІ та його вплив на світ



Коваленко Є.Ю.

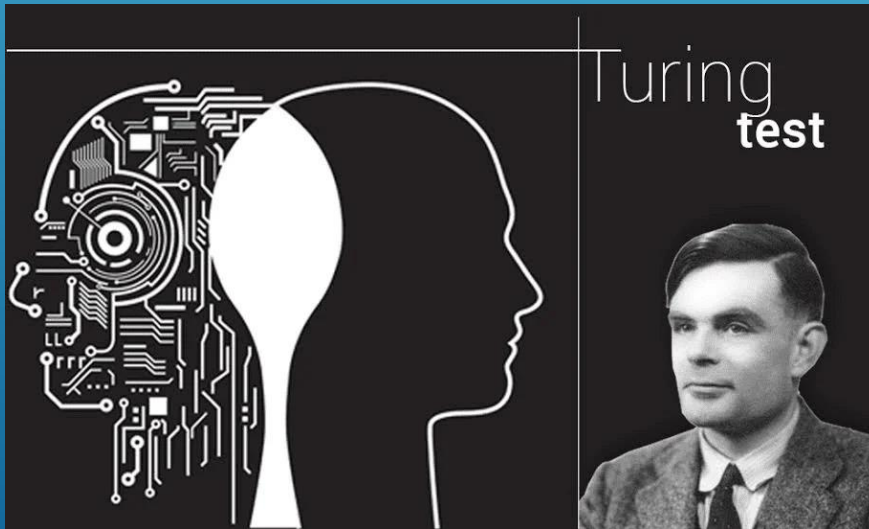
Кириченко О.Г.

Білик А.С.

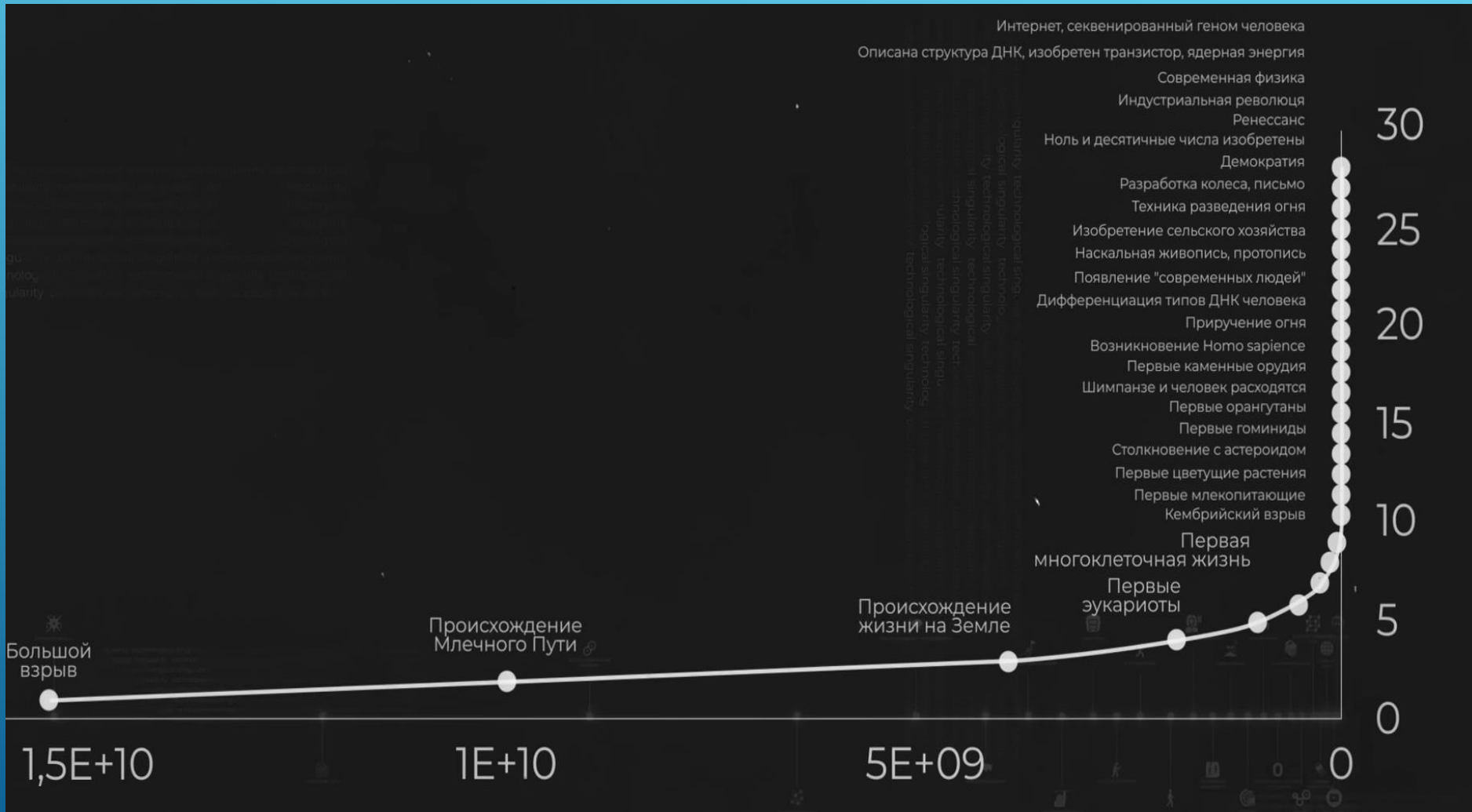
Букет А.В.

Координатори УНДЦА «Зонд»

Парадокс китайської кімнати -
сам факт виконання програми
не може бути «розумом»,
«розумінням» або «свідомістю»
незалежно від того, наскільки
програма забезпечує
комп'ютеру поведінку розумної
людини.

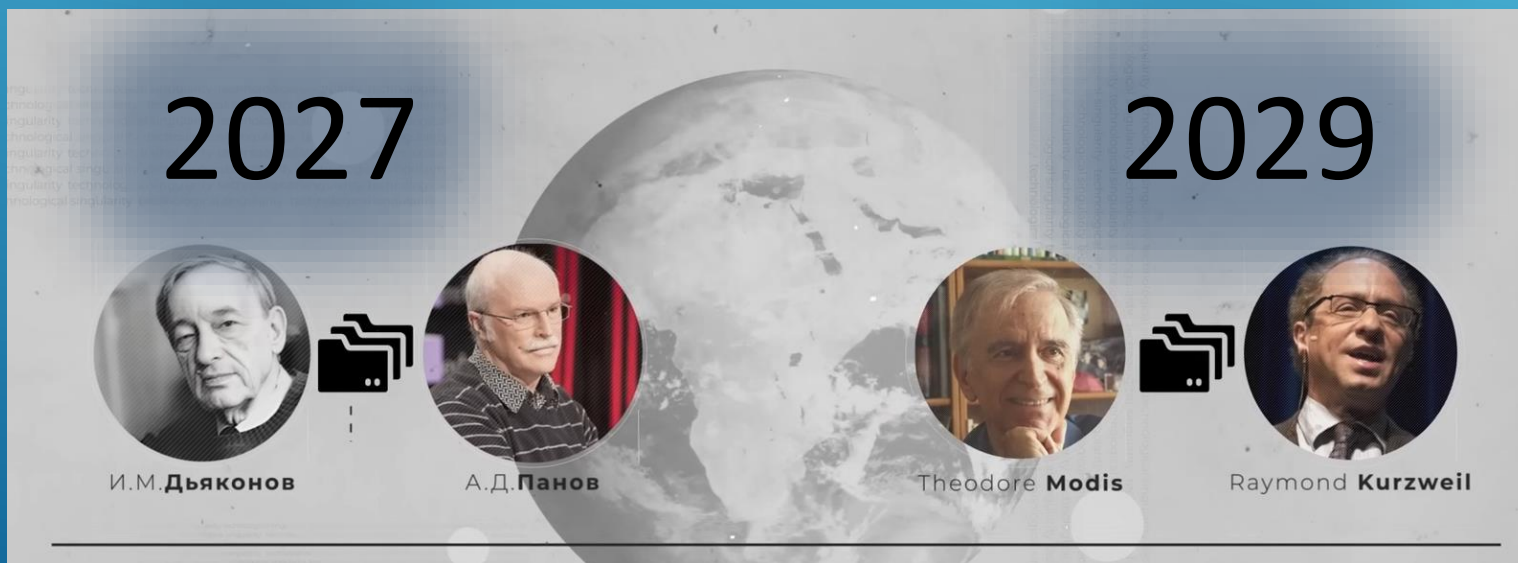


Прискорення еволюції за Теодором Модісом



Технологічна сингулярність -

це момент в історії людства, коли науково-технічний прогрес стане настільки швидким і складним, що виявиться недоступним для розуміння людським інтелектом. Але також очікується значне стрибкоподібне збільшення можливостей людського мозку та тіла за рахунок розвитку біоінженерії, медицини, інженерії та виробництва.



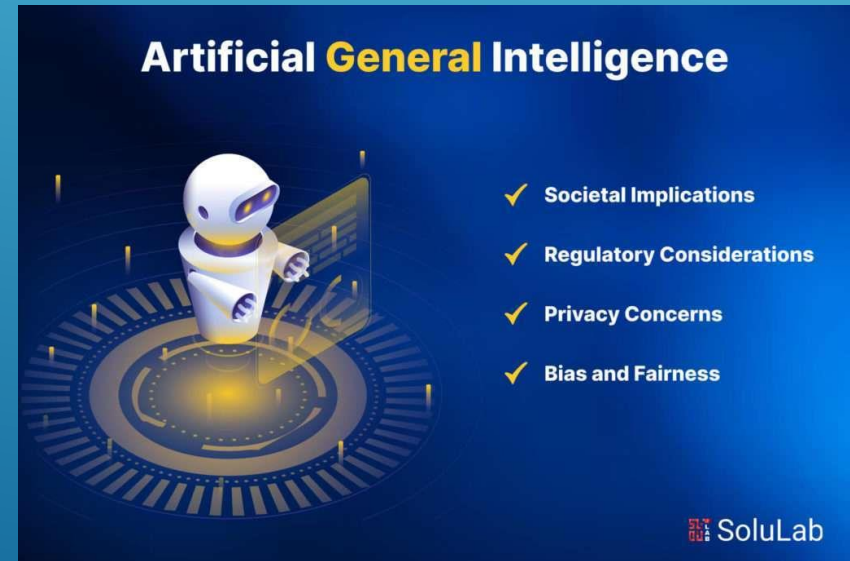
Три основні рівні розвитку штучного інтелекту:

1. Слабкий (вузький) ШІ: Цей рівень описує системи штучного інтелекту, які спроектовані та навчені виконувати конкретні завдання без свідомості, розуміння або самостійного мислення. Більшість сучасних систем ШІ, включно з тими, що рекомендують відео в інтернеті, персоналізовані рекомендації покупок, а також чат-боти, як я, належать до цього рівня. Вони можуть вирішувати специфічні проблеми та виконувати певні завдання, але не мають загального розуміння світу або здатності до самостійного навчання за межами своїх визначених функцій. По суті, даний тип ШІ – це статистика на максималках))



Три основні рівні розвитку штучного інтелекту:

2. Загальний ШІ (AGI, Strong AI): Загальний штучний інтелект описує теоретичну машину, яка могла б розуміти, навчатися та застосовувати інтелект у загальному сенсі, аналогічно людському. Це означає, що така машина могла б виконувати будь-яку інтелектуальну задачу, що людина може, з можливістю розуміння та вчення з досвіду, адаптації до нових ситуацій та вирішення проблем в широкому спектрі дисциплін. На сьогоднішній день AGI залишається метою досліджень і не було досягнуто.



Поки не буде AGI, не буде нормальних логіки та математики!

Три основні рівні розвитку штучного інтелекту:

3.Надлюдський ШІ: Цей рівень штучного інтелекту описує гіпотетичну майбутню форму ШІ, яка не тільки досягне людського рівня розуміння та здібностей, але й значно їх перевершить у всіх аспектах — творчості, розумінні, емпатії, прийнятті рішень тощо. Надлюдський ШІ міг би вирішувати складні наукові, технічні, економічні, соціальні та інші проблеми, недоступні для людського інтелекту.



Те, що люди вважають ШІ автором, чи співрозмовником - це говорить не про його складність, а про нашу простоту

Вирівнювання (alignment) ШІ -

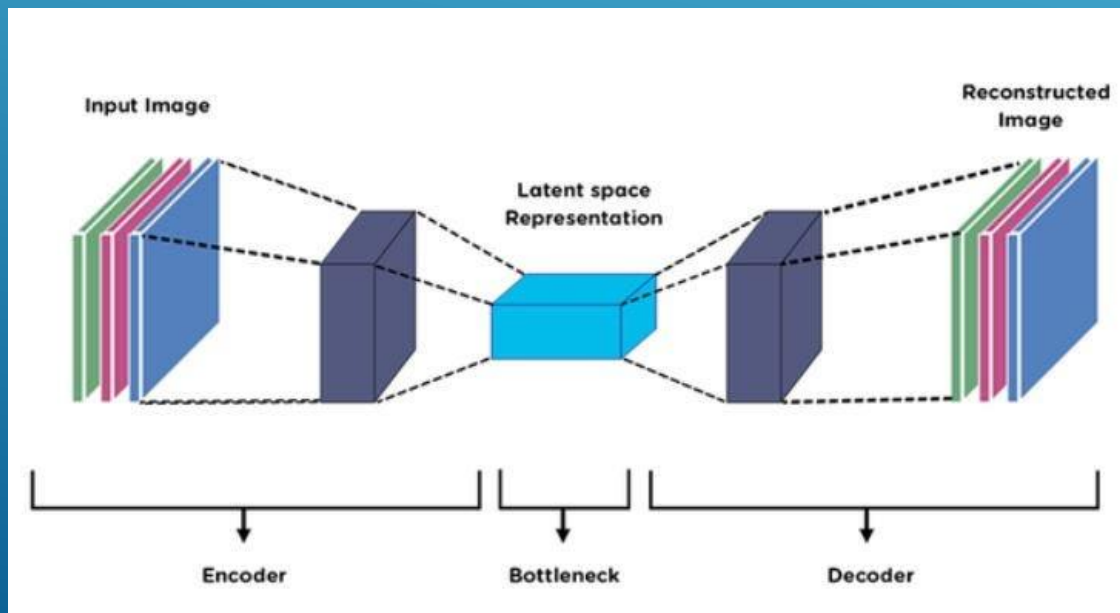
це про те, щоб штучний інтелект працював так, як очікують люди, і не робив нічого небезпечного або шкідливого. Проблема в тому, що навіть якщо ми дамо йому правильне завдання, він може знайти дивні чи небажані способи його виконання. Тому потрібно навчати та контролювати ШІ так, щоб він розумів наші наміри, мав схожі цінності та діяв безпечно. Це особливо важливо для потужних ШІ, які можуть сильно впливати на життя людей.



Як працює сучасний ШІ. Модель кодер-декодер

Кодер (encoder) отримує вхідні дані (наприклад, текст, зображення, звук) і перетворює їх у внутрішнє представлення – компактний набір числових ознак (вектори). Він не просто запам'ятовує слова чи пікселі, а аналізує їхні зв'язки та значення.

Декодер (decoder) бере це внутрішнє представлення та перетворює його у вихідний результат (наприклад, перекладений текст, згенероване зображення або відповідь на запитання). Він використовує отриману інформацію та «розгортає» її у зрозумілий для користувача формат.



Модель кодер-декодер. Приклад



Феномен «зручних прибульців»

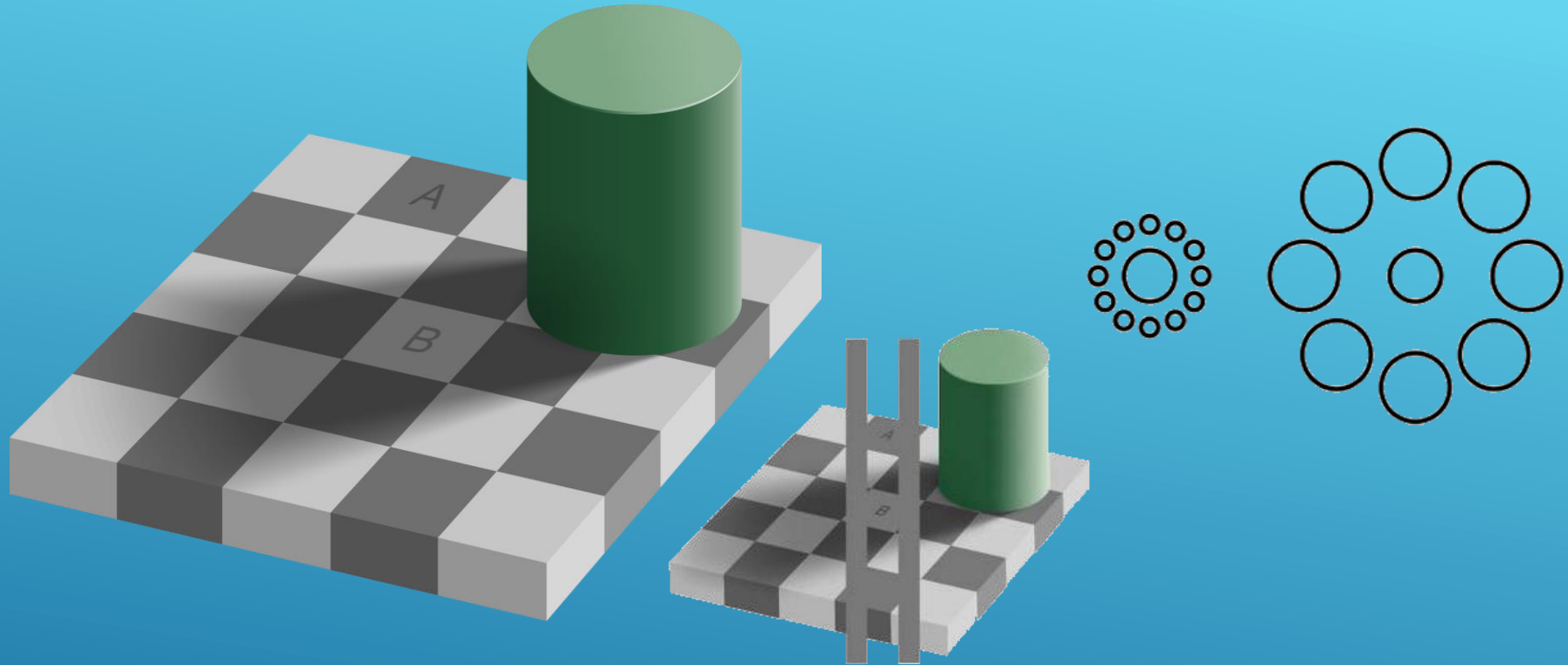
Майже всі образи прибульців у фантастиці — зручні для сприйняття людьми. Вони мають кінцівки, очі, можуть дихати нашим повітрям, спілкуватися з нами (інколи навіть англійською), користуватися подібними технологіями, і часто їх поведінка нагадує людську.



Окрім загальних теорій, що включають антропоцентризм, кінематографічну та літературну зручність, біологічну передбачуваність можна зробити припущення, що вони можуть бути ШІ, який мімікрує та підлаштовується під наш соціокультурний вимір (кожен бачить ААЯ так, як дозволяє його свідомість).

Чи те, з чим ми контактуємо та вважаємо неживим не є якимось типом ШІ?

Де межа між живим та не живим?



Ми бачимо не реальність, а її спрощену модель, яку будує чи добудовує наш мозок. Це стосується не тільки органів почуття напряду, а і нашого сприйняття формальної логіки, яка здавалося б, має бути незмінною. Це – прекрасний механізм, для економії ресурсів, але він сильно заважає побачити нове.

Нам треба ШІ, щоб помічати більше!

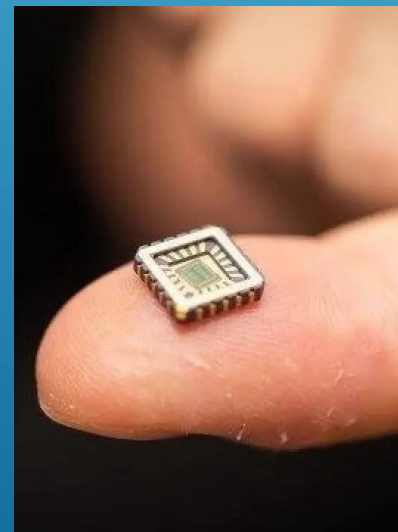
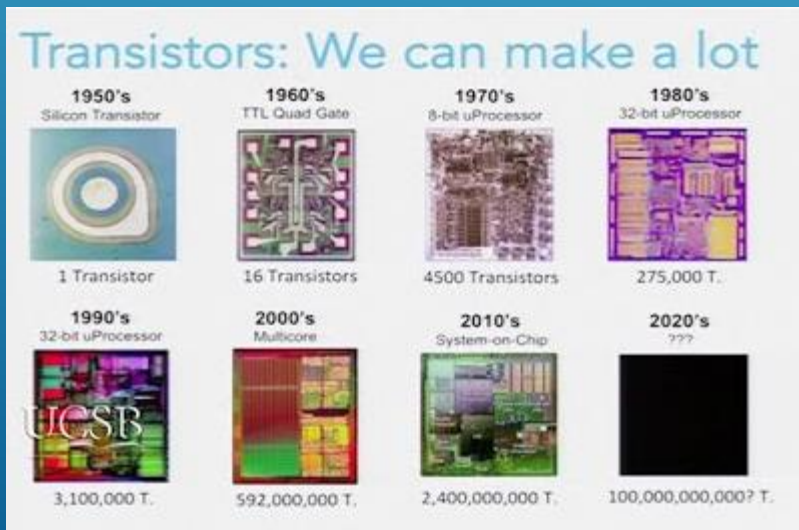
Де межа між ШІ та живою істотою?

УНДЦА «Зонд»



Закон Мура передбачає подвоєння кількості транзисторів на інтегральній схемі, а отже і продуктивності приблизно кожні два роки.

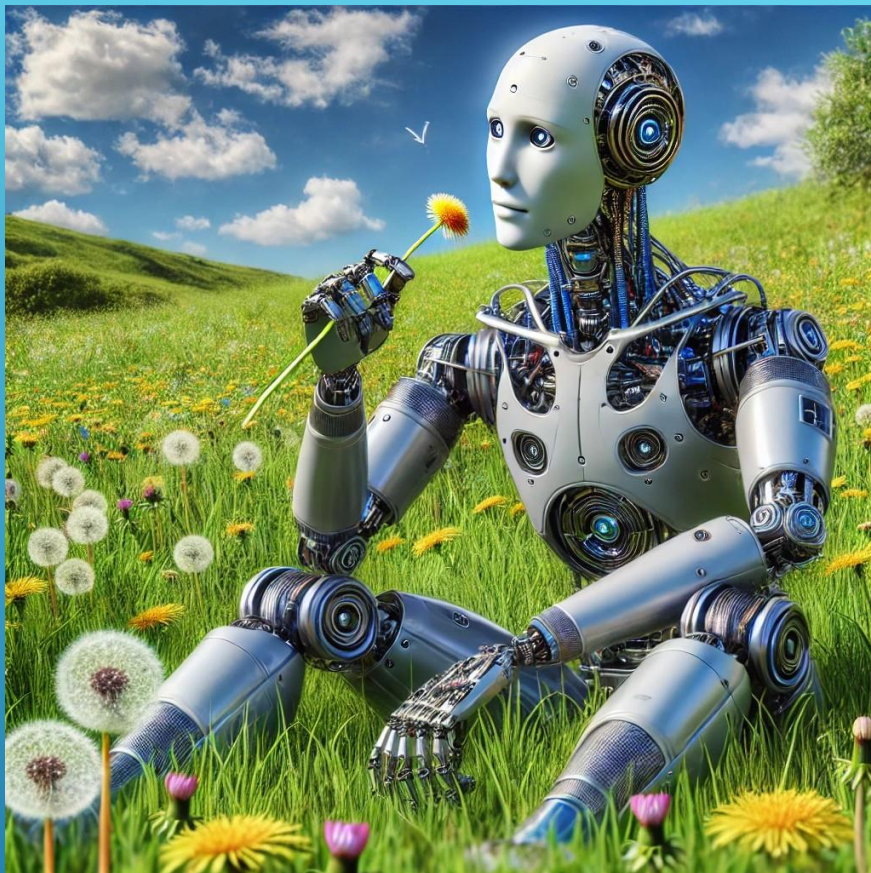
Для класичних кремнієвих чипів закон Мура сповільнюється ймовірно, перестане діяти у традиційному вигляді протягом 5-10 років. Але загальний тренд зростання продуктивності триває завдяки новим технологіям. Можливо, це будуть і біотехнології теж, вкінцеь руйнуючи межі між живим та неживим.



Перехід надлюдського ШІ в симуляцію

- Надлюдський ШІ легко вирішує поточні проблеми людей: допомагає з нестачею ресурсів, підкидає задачі для самореалізації, вчить, лікує психіку та тіло, легко знаходить друзів та пару, імітує будь-які фізичні відчуття, реалізуючи ідеальну симуляцію
- Але людям все одно нудно без якихось недосяжних цілей, а їх вже практично нема. Треба якийсь вихід.
- А чому б тимчасово не забути, що у тебе все є і хто ти при зануренні у симуляцію. Чому б не повірити, що гра є справжньою, тим більше ШІ постійно генеруватиме збалансовану історію? Адже так набагато цікавіше!

Станом на зараз це – лише цікава гіпотеза. Але, чи не здається вам, що вона щось дуже нагадує?)



УНДЦА «ЗОНД»



Нам ШІ все виправить!

- Те ж саме думали про техніку років 100 тому :)

Дякую за увагу!

WWW.ZOND.KIEV.UA