

Вибір майбутнього: AGI чи ASI? Штучний інтелект на роздоріжжі між проривом і відповідальністю



*Юліан Бедь, Роман Чийпеш
Карпатський університет імені Августина
Волошина*

м. Ужгород, 17 липня 2025 року

Аналітично-наукова стаття про виклики розвитку AGI та ASI в умовах глобальної конкуренції та етичної відповідальності

Технологія, що змінює цивілізацію

Штучний інтелект (ШІ) більше не є просто технологічною інновацією — він стає визначальним викликом цивілізаційного масштабу. Станом на 17 липня 2025 року світ стоїть перед вибором: чи йти етапним шляхом розвитку до AGI (ей-джі-ай, англ. Artificial General Intelligence - артіфішал дженерал інтелідженс) — штучного загального інтелекту, чи дозволити стрімкий стрибок до ASI (ей-ес-ай, англ. Artificial Superintelligence - артіфішал сьюперінтелідженс) — штучного надінтелекту.

Цей вибір визначатиме не тільки майбутнє технологій, але й саму долю людства.

Глосарій основних понять

- AGI (ей-джі-ай) — гіпотетичний інтелект, здатний самостійно вирішувати широкий спектр завдань, як і людина: логіка, навчання, творчість, адаптація.
- ASI (ей-ес-ай) — гіпотетична форма інтелекту, яка значно перевершує людський у всіх сферах: від стратегії до етики.
- Інтелектуальна експлозія — моментальний стрибок від AGI до ASI внаслідок самопокращення і масштабування.

- Alignment (елайнмент) — узгодження цілей і дій ШІ із людськими моральними та етичними цінностями.
- Сценарій прямого стрибка — перехід до ASI без етапу AGI, що значно підвищує ризики втрати контролю.
- Контрольована еволюція — поетапний розвиток ШІ із запровадженням систем тестування, регулювання і моральної перевірки.

Оновлення станом на липень 2025 року

- Future of Life Institute (Ф'ючер оф лайф інстїт्यूт - Інститут майбутнього життя) опублікував звіт (липень 2025), згідно з яким жодна з провідних AI-компаній (OpenAI – Оупен ей-ай, Anthropic - Антропік, DeepMind - ДїпМайнд тощо) не отримала оцінку вищу за «С» за рівень безпеки у розробці AGI.
- Andrew Ng (Ендрю Енг) наголосив: AGI — це спекулятивна концепція, яка відволікає від реальних задач ШІ.
- Forbes (Форбс) підтвердив: мультиагентна AGI-система вже перевершила людей у низці спеціалізованих завдань. Це може сигналізувати про швидкий стрибок до ASI.
- New York Post (Нью-Йорк Пост) попереджає про можливість створення ASI протягом 10 років. Існує серйозний дефіцит міжнародного контролю.
- AI Act EU (Ей-Ай Акт ЄЮ - регуляторний акт Європейського Союзу щодо штучного інтелекту) — перші положення щодо регуляції «загальних моделей ШІ» набирають чинності вже в серпні 2025 року.
- Рамкова конвенція про штучний інтелект, права людини, демократію та верховенство права Ради Європи від 17 травня 2024 року (підписана й Україною) — перша міжнародна угода, яка юридично зобов'язує дотримання прав людини при розробці ШІ.

Стратегічна дилема: поступ чи стрибок?

Сценарій прямого стрибка до ASI

У разі, якщо деякі системи ШІ досягнуть здатності до самонавчання, масштабування й адаптації без етапу AGI, людство опиниться перед неконтрольованим надінтелектом. Слабкість чинних регуляторних систем та відсутність інституційної відповідальності загострює загрозу.

Контрольована еволюція

Альтернативою є етапний підхід: від вузького ШІ до AGI — з обов'язковим впровадженням alignment (елайнмент – узгодження цілей, рішень і дій штучного інтелекту з моральними, етичними та правовими цінностями людини), а далі — до ASI лише за умови наявності глобального нагляду й етичної перевірки.

Alignment: чи можливо навчити ШІ діяти на користь людини?

Проблема полягає в тому, що навіть надзвичайно розумна система, не маючи вбудованих цінностей, може діяти логічно, але неетично. Поточні технології alignment включають:

- Human Feedback RL (г'юман фідбек ер-ел - навчання з підкріпленням за допомогою зворотного зв'язку від людини);
- Inverse Reinforcement Learning (інверс рїнфорсмент льорнінг - зворотне навчання з підкріпленням);
- Етичні інтерпретатори запитів;
- Архітектурні обмеження, як у GPT-моделях (джі-пі-ті) від OpenAI.

Чому діяти треба вже зараз?

- AGI (ей-джи-ай) може бути створений вже за 3–5 років, а ASI (ей-ес-ай) — за 10–15.
- Відсутність глобальних угод загрожує транснаціональною нестабільністю.
- Етичний аудит та прозорість мають стати обов'язковими вимогами.

Прогноз і рекомендації

1. Заснувати міжнародний центр аудиту AGI/ASI з правом екстреного призупинення проєктів.
2. Розробити “червоні лінії” для застосування ASI — особливо у військовій сфері.
3. Zobov'yazati kompanii publikuvati protokoli bezpeki alignment (елайнмент — узгодження цілей, рішень і поведінки штучного інтелекту з цінностями, етикою та інтересами людини).
4. Забезпечити навчання та участь громадян у прийнятті рішень щодо ШІ.

Висновок

AGI та ASI — це не лише етапи технічної еволюції, а й тест на моральну зрілість людства. Вибір між "етичним прогресом" і "неконтрольованим проривом" — це вибір між цивілізаційним оновленням і потенційним самознищенням.

Істинний прорив у ШІ — не в тому, щоб створити розумнішого за нас, а в тому, щоб зробити його етичним, відповідальним і гуманним.

Джерела

1. Guardian. AI firms 'unprepared' for dangers of building human-level systems, report warns. 17 липня 2025 р. [Електронний ресурс]. – Режим доступу: <https://www.theguardian.com/technology/2025/jul/17/ai-firms-unprepared-for-dangers-of-building-human-level-systems-report-warns> – Назва з екрана. – Дата звернення: 17.07.2025.
2. Business Insider. Andrew Ng: AGI is overhyped. 11 липня 2025 р. [Електронний ресурс]. – Режим доступу: <https://www.businessinsider.com/google-brain-founder-andrew-ng-agi-is-overhyped-ус-2025-7> – Назва з екрана. – Дата звернення: 17.07.2025.
3. Forbes. AGI to ASI Pathway: Future Forecasting the Rise of AI Superintelligence. 9 липня 2025 р. [Електронний ресурс]. – Режим доступу: <https://www.forbes.com/sites/lanceeliot/2025/07/09/future-forecasting-the-agi-to-asi-pathway-giving-rise-to-ai-superintelligence> – Назва з екрана. – Дата звернення: 17.07.2025.
4. New York Post. We need guardrails for artificial superintelligence before it's too late. 8 червня 2025 р. [Електронний ресурс]. – Режим доступу: <https://nypost.com/2025/06/08/opinion/we-need-guardrails-for-artificial-superintelligence-before-its-too-late> – Назва з екрана. – Дата звернення: 17.07.2025.
5. European Commission. AI Act – Regulatory framework on Artificial Intelligence. [Електронний ресурс]. – Режим доступу: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> – Назва з екрана. – Дата звернення: 17.07.2025.
6. Council of Europe. Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS 225)/Рамкова конвенція Ради Європи про штучний інтелект, права людини, демократію та верховенство права. 17 Мау 2024. [Електронний ресурс]. – Режим доступу: офіційний PDF-документ Ради Європи (<https://rm.coe.int/1680afae3c>). – Назва з екрана. – Дата звернення:

17.07.2025.