

Анотована історія сучасного ШІ та глибокого навчання

Від Лейбніца та Івахненка до Трансформерів

Олександр Ковальов, IM-51mn

«Київський політехнічний інститут імені Ігоря Сікорського»
Група IM-51mn

25 листопада 2025 р.

План виступу

- 1 Фундамент (1676-1958)
- 2 Київський прорив (1965)
- 3 Backpropagation та Архітектури
- 4 LSTM та GPU
- 5 Висновки

Міф про 2012 рік

Поширений міф

Штучний інтелект народився у 2012 році з появою AlexNet та потужних GPU.

Реальність:

- Математика була готова у XVII столітті.
- Перші нейромережі працювали у XIX столітті.
- Глибоке навчання винайшли у Києві в 1960-х.

Лейбніц і Ланцюгове правило (1676)

Основа навчання сучасних мереж — **Backpropagation**. Математично це просто ланцюгове правило диференціювання:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial w} \quad (1)$$

- **Готфрід Вільгельм Лейбніц (1676):** Сформулював це правило.
- Без цього ми б не могли вирішити *Credit Assignment Problem* (хто винен у помилці?).

Перцептрон Розенблатта (1958)

Френк Розенблатт, Mark I:

- Перша фізична машина, що навчалася.
- Ейфорія в New York Times.

Проблема: Це було **Shallow Learning**.

- Навчався лише останній шар.
- Не міг вирішити задачу XOR (Мінський, 1969).

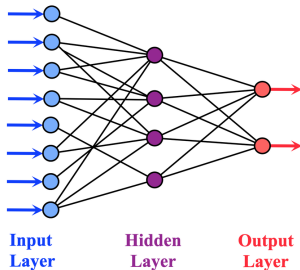


Рис. 1: Схема: Навчається тільки вихід

Народження Глибокого Навчання в Києві

Поки Захід переживав "Зиму ШІ у Києві (Інститут кібернетики) відбувалася революція.

Олексій Івахненко та Валентин Лапа (1965)

Розробили **МГУА (GMDH)** — Метод групового врахування аргументів.

- Перший алгоритм для **багатошарових** мереж.
- Мережа росла сама (самоорганізація).
- Функція активації — поліном:

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2 \quad (2)$$

Світовий рекорд 1971 року

Досягнення Івахненка (1971):

- Працююча глибока мережа.
- **8 шарів** обробки інформації!
- У той час на Заході ледве працювали з 1-2 шарами.

Чому забули?

- 1 Публікації російською/українською.
- 2 Відсутність Backprop для фіксованих структур.
- 3 Слабкі комп'ютери 70-х.

Справжня історія Backpropagation

Міф: Винайдено у 1986 (Хінтон).

Реальність: Це "перевідкриття".

- **1960-1961 (Келлі, Брайсон):** Використання для керування ракетами (аналітичний градієнт).
- **1970 (Сеппо Ліннаінмаа):** *Reverse Mode of Automatic Differentiation*.

Саме Ліннаінмаа показав, як рахувати похідні ефективно:

$$\text{Cost}(\nabla f) \approx K \cdot \text{Cost}(f) \quad (3)$$

Еволюція: Неокогнітрон (1979)

Куніхіко Фукусіма (1979):

- Предтеча сучасних CNN.
- Натхнення: зорова кора котів (Хьюбел і Візел).

Архітектура:

- 1 S-шари (Simple) → Згортка.
- 2 C-шари (Complex) → Пулінг.

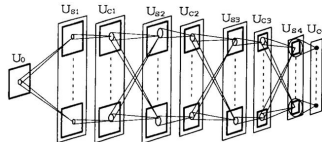


Рис. 2: Неокогнітрон

Проблема зникаючого градієнта (1991)

У глибоких мережах (особливо RNN) градієнт множиться багато разів:

$$\frac{\partial E}{\partial w} \propto \prod_{t=1}^T w_{rec} \quad (4)$$

Якщо вага < 1 , градієнт прямує до 0. Мережа перестає вчитися.

Рішення: LSTM (Хохрайтер і Шмідхубер, 1997)

- Constant Error Carousel (CEC).
- Градієнт **додається**, а не множиться.

Ера GPU: DanNet (2011) проти AlexNet (2012)

Революція почалася не з алгоритмів, а зі заліза (GPU).

- **2010:** Ден Чірешан (команда Шмідхубера) портує глибокі мережі на CUDA (50x прискорення).
- **2011 (DanNet):** Перша перемога над людиною (Traffic Sign Recognition, 0.56% помилки).
- **2012 (AlexNet):** Перемога на ImageNet. Це популяризація підходу Чірешана на більших даних.

Сучасність і Висновки

Трансформери (2017):

- Механізм *Attention* має коріння в концепції «Fast Weights» (Шмідхубер, 1991).
- ChatGPT — це масштабування ідей минулого століття.

Головний висновок

Історія ШІ циклічна. Ми стоїмо на плечах гігантів: Лейбніца, Івахненка, Ліннаїнмаа та Фукусіми.

Дякую за увагу!

Запитання?