

Сучасні підходи використання нейронних мереж в аналізі текстів

Драговоз Даміан,
КН-3,
2025.

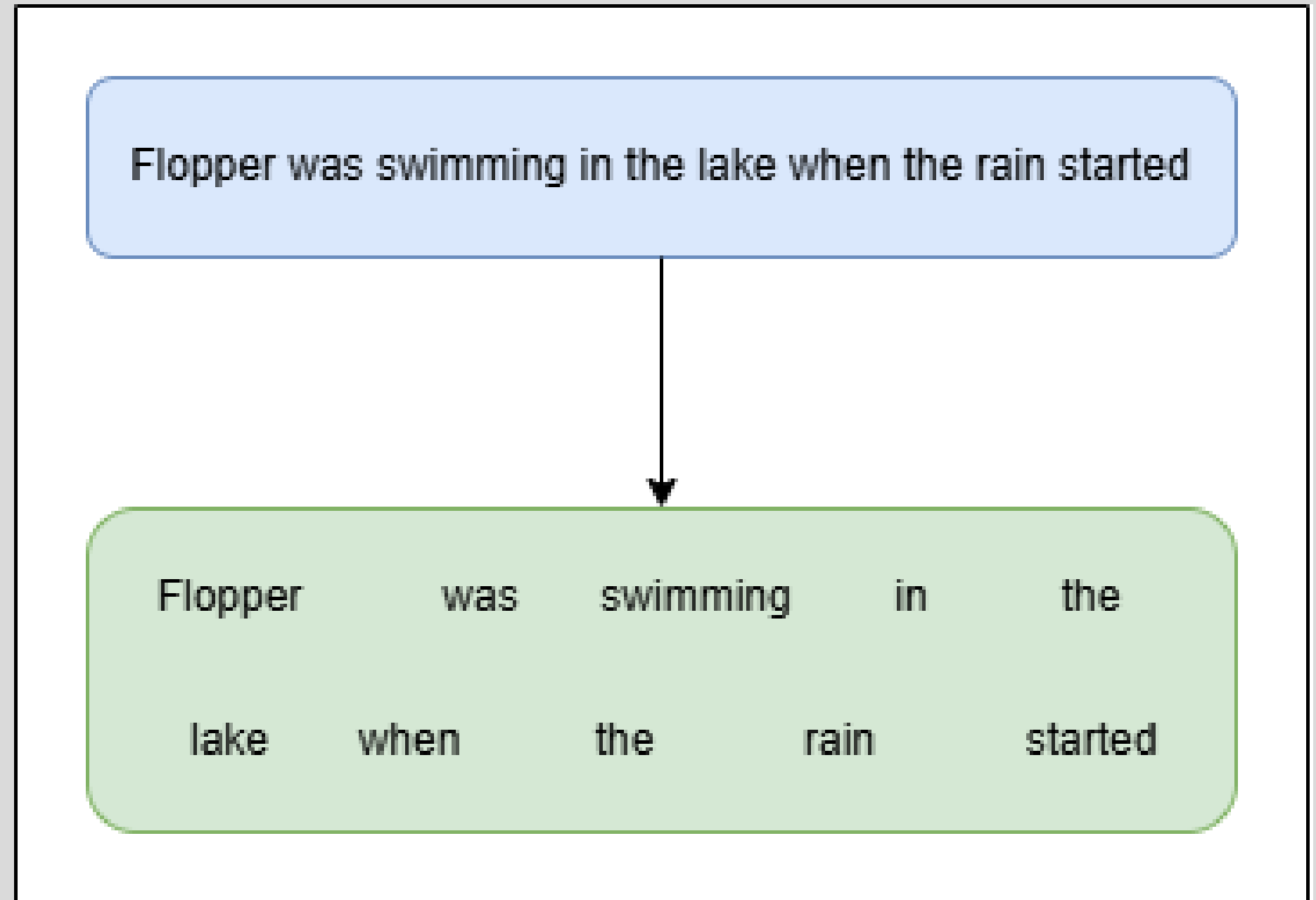
Керівник:
Глибовець Андрій Миколайович,
Доктор технічних наук, декан.

ВСТУП

- Штучний інтелект зазнав значного зростання завдяки проривам, таким як моделі GPT
- Ця презентація досліджує фундаментальні концепції NLP з акцентом на нейронні підходи
- Ми розглянемо прогрес від базової токенизації до архітектур трансформерів

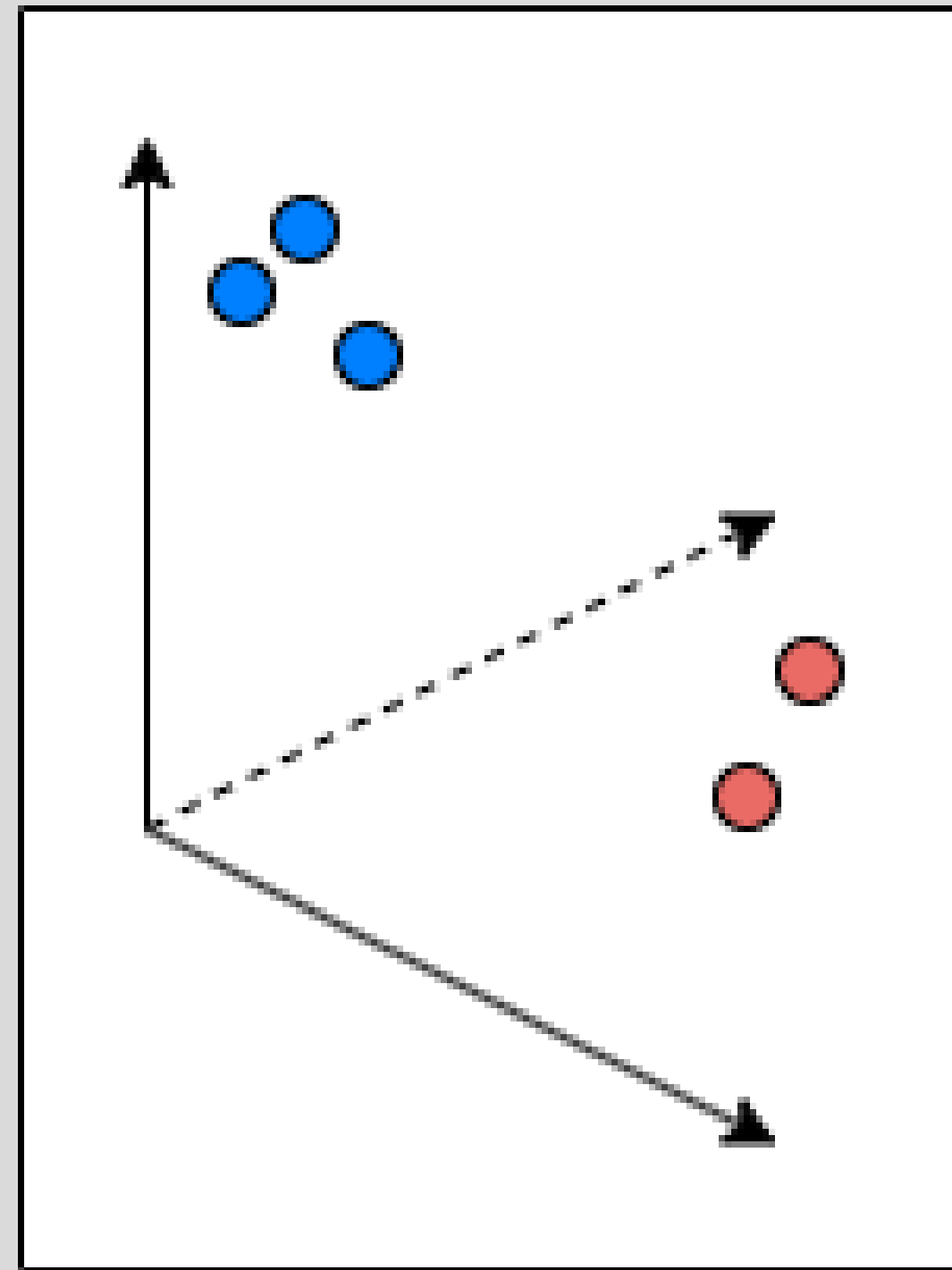
ТОКЕНІЗАЦІЯ

- Перший крок в обробці тексту
- Byte-Pair Encoding



ПРЕДСТАВЛЕННЯ СЛІВ

- Оцінки створені людиною
- Term-term матриця
- Word Embeddings



НЕЙРОННІ МЕРЕЖІ

- Feed-forward
- Recurent
- Long-Short Term Memory
- Transformer

ВЛАСНІ ЕКСПЕРИМЕНТИ

- Реалізація BPE токенизації та тестування ефективності для різних мов
- Побудова RNN та LSTM моделей для генерації тексту
- Створення власних векторних представлень слів та візуалізація семантичних зв'язків
- Спостереження покращення узгодженості зі збільшенням складності архітектури

ПРИКЛАДИ РЕЗУЛЬТАТІВ (EMBEDDINGS)

```
[('and', 1.0000001192092896),  
 ('but', 0.8101518154144287),  
 ('so', 0.7278461456298828),  
 ('so.', 0.7171671390533447),  
 ('yeah.', 0.7127646207809448),  
 ('you...', 0.7123894095420837),  
 ('nowhere', 0.7083984613418579),  
 ('vand', 0.7060924768447876),  
 ('you', 0.6926586031913757),  
 ('well.', 0.692141056060791)]
```

```
[('брат', 0.9999997615814209),  
 ('брат.', 0.9686229228973389),  
 ('братами', 0.9476014971733093),  
 ('братів', 0.9450318813323975),  
 ('братові', 0.9435192942619324),  
 ('трафить', 0.8901544213294983),  
 ('котра', 0.8753548264503479),  
 ('дратували', 0.8727647066116333),  
 ('дратувати', 0.8700396418571472),  
 ('завтра', 0.8674534559249878)]
```

ПРИКЛАДИ РЕЗУЛТАТИВ (RNN GENERATION)

harry spun back through a barful as the snake

harry woke them. he looked down, down at his

‘i’m ’ npt tp try, potter?

‘oh, i’m glad to be chess of the teachers championshit the
gamekeeper. And i’ll lose you loads,” said hagrid, patting his
crossbow.

ПРИКЛАДИ РЕЗУЛЬТАТІВ (LSTM)

harry had to see it. But she had been in one of a very deal, and
he was sure, but the dambledore was going for the ministry

“well, you know,” he asked her voice to the gargoies

НЕДОЛІКИ ТА МАЙБУТНЯ РОБОТА

1. Відсутність тестувань архітектур для серйозних датасетів

2. Не достатньо проведеного досліджень архітектури

Трансформера

1. Дослідження складніших архітектур

2. Тестування архітектур для ширшого списку NLP проблем

Дякую за увагу