

## АНАЛІЗ ЕФЕКТИВНОСТІ МОДЕЛЕЙ ЧАТ-БОТІВ НА БАЗІ АРХІТЕКТУР RNN ТА LSTM

Бердієва С. Х., Статкус А. В.

*Національний технічний університет  
«Харківський політехнічний інститут», м. Харків*

У контексті стрімкого розвитку інформаційних технологій особливу увагу привертають інтелектуальні системи обробки природної мови, зокрема чат-боти (ЧБ). Інтеграція алгоритмів глибинного навчання (ГН) в такі системи дозволяє значно підвищити якість діалогової взаємодії.

*Метою доповіді* є дослідження застосування ГН у чат-ботах на основі поширених архітектурних моделей рекурентних нейронних мереж.

Моделі ГН в ЧБ використовуються для реалізації віртуальних помічників, автоматизованих систем підтримки клієнтів, освітніх платформ, систем первинного діагностування в медицині, а також як частина інтерфейсів користувача в складних інформаційних системах. Серед ключових науково-технічних викликів — забезпечення контекстної узгодженості відповідей, уникнення генерації семантично некоректних чи тривіальних реплік, вибір архітектурної моделі, що відповідає конкретним умовам використання, та забезпечення стійкості моделей до «шуму» в даних і нових типів запитів.

В якості методу дослідження впливу архітектурних особливостей рекурентних нейронних мереж (НМ) на якість роботи діалогової системи обрано експериментальний порівняльний аналіз ефективності ЧБ на основі архітектур RNN та LSTM.

Два відповідних прототипи ЧБ реалізовані на платформі PyTorch, яка порівняно з TensorFlow має простіший і більш інтуїтивний API і забезпечує динамічне обчислення графу, що спрощує налагодження та експерименти з архітектурами НМ і особливо корисне при розробці ЧБ. Для забезпечення сумісності результатів обидві моделі (RNN та LSTM) навчались на однаковому корпусі діалогів, який сформовано як JSON-файл з чіткою структурою "інтент – приклади – відповіді". Діалоги мали середню довжину у 5–8 слів і включали лексичне різноманіття, що дозволяло перевірити здатність моделей до узагальнення і коректної класифікації інтенів. Ефективність ЧБ визначалася об'єктивними метриками (perplexity, BLEU) і експертними оцінками семантичної повноти, логічної узгодженості та релевантності відповідей. З метою оцінки генералізаційних можливостей моделей також було враховано їх реакцію на нові запити, раніше не представлені в навчальному наборі.

RNN-модель показала гарні результати для коротких запитів, але поступалася при обробці складніших конструкцій або варіацій формулювань. LSTM-модель краще справлялася з довшими запитами і варіативністю мови завдяки здатності зберігати контекст на довших часових відрізках.

У разі розгортання ЧБ для реального інтернет-магазину, коли користувачі формулюють запити у довільній неструктурованій формі, рекомендується LSTM-архітектура, яка показала вищу адаптивність, точність і стійкість до шумів. Проте, якщо мова йде про обмежене середовище з чітко визначеними командами або умови обмежених обчислювальних ресурсів, – можлива реалізація простішого рішення на базі RNN.