

О.А. ПАСТУХ

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна, e-mail: ol_pas@tntu.edu.ua.

В.В. ЯЦИШИН

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна, e-mail: yacyshyn@tntu.edu.ua.

О.М. ЗАДВОРНИЙ

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна, e-mail: zadvorni.alex16@gmail.com.

ТЕХНОЛОГІЯ ГЕНЕРАЦІЇ ТА ОЦІНЮВАННЯ ЯКОСТІ ПРОГРАМНОГО КОДУ ВЕЛИКИМИ МОВНИМИ МОДЕЛЯМИ

Анотація. Запропоновано технологію генерації програмного коду на основі zero-shot prompting та метрику оцінювання його інтегральної якості, яка включає чотири групи метрик: повну впевненість моделі, семантичну якість, структурну цілісність та динамічне виконання коду. Проведено емпіричне дослідження двох типів структури запитів. Запропоновано метрику, що дає змогу отримувати відтворюваний, комплексний показник якості, який поєднує функціональну коректність із показниками супроводжуваності та надійності коду. Наведено результати, що формують основу для подальшої стандартизації промпт-інженерії та розвитку підходів до об'єктивного оцінювання генерації програмного коду великими мовними моделями у реальних інформаційних системах.

Ключові слова: великі мовні моделі, генерація програмного коду, оцінювання якості коду, zero-shot prompting, промпт-інженерія, метрика інтегральної якості.

DOI 10.34229/KCA2522-9664.26.3.2

ВСТУП

Поява великих мовних моделей (ВММ) спричинила значні зміни у сучасній інженерії програмного забезпечення. Моделі від OpenAI, DeepSeek, Qwen та інших компаній тепер здатні генерувати складний, функціональний код з описів природною мовою, таких як user-stories або технічні вимоги. Це досягнення може кардинально прискорити цикли розроблення, автоматизувати створення шаблонного коду та підвищити продуктивність розробників.

Оскільки інтеграція ВММ у комп'ютерні системи та робочі процеси інформаційних технологій стає реальністю, фокус зміщується з простої генерації коду моделями на отримання найоптимальнішої версії коду. Запит природною мовою є основним інтерфейсом для цієї взаємодії, що робить промпт-інженерію вагомим фактором у продуктивності моделі. Однак нерозв'язною проблемою залишається структура запиту, сформульованого природною мовою, та його вплив на якість згенерованого результату, зокрема, як структура запиту до ВММ у вигляді згрупованого набору вимог за функціональними модулями впливає на якість генерування програмного коду порівняно із структурою запиту, що містить лише типовий набір вимог.

Для оцінювання впливу структури запиту потрібно забезпечити вимірюваність якості програмного коду за багатьма метриками, а не лише за функціональною коректністю. У наявних стандартах часто зроблено акцент на алгоритмічну правильність [1–3], що є недостатнім для оцінювання застосовності програмного коду, призначеного для реальних інформаційних систем. Отже, поза увагою залишаються кількісні та якісні атрибути, зокрема:

- супроводжуваність, дотримання стилістичних конвенцій та безпечність типів;
- працездатність, що реалізує набір взаємопов'язаних вимог;
- впевненість ВММ у згенерованому нею програмному кодї.