

УДК 004.89

*Поздняков О. А., аспірант,
Пархоменко А. В., к. т. н., доцент,
Національний університет «Запорізька політехніка»*

ОЦІНЮВАННЯ НЕВИЗНАЧЕНОСТІ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ПРИ ГЕНЕРАЦІЇ МІГРОВАНОГО КОРИСТУВАЦЬКОГО АВАР-КОДУ

Інтелектуальний реінжиніринг складних комп'ютерних систем (КС), зокрема міграція користувацького АВАР-коду при переході з застарілої версії SAP ERP 6.0 на нову версію SAP S/4HANA, є одним із ключових завдань цифрової трансформації сучасних підприємств [1].

Дослідження [2, 3] показали ефективність використання методів та засобів генеративного штучного інтелекту для автоматизації цього процесу. Разом з тим, залишається невирішеним питання довіри до результатів, оскільки велика мовна модель (LLM) може генерувати некоректний код без явних ознак помилки (hallucinations), що є критичною проблемою у корпоративних системах. Проблематика оцінювання невизначеності (Uncertainty Quantification, UQ) для LLM є предметом активних досліджень [4, 5], проте її застосування в задачах міграції користувацького коду залишається недостатньо вивченим, що обумовлює актуальність цієї роботи.

Метою роботи є розробка методу оцінювання невизначеності LLM при генерації мігрованого користувацького АВАР-коду для підвищення ефективності процесу інтелектуального реінжинірингу КС.

Розроблений метод складається з трьох етапів.

Перший етап передбачає введення метрики Confidence Score (CS) на основі ентропії розподілу токенів при генерації:

$$CS = 1 - \frac{H(p)}{\log|V|},$$
 де $H(p)$ – ентропія розподілу ймовірностей наступного токена, $|V|$ – розмір словника моделі. Висока ентропія сигналізує про невизначеність моделі під час генерації коду.

Другий етап базується на стохастичному семпльованні (Stochastic Sampling, SS): для одного вхідного фрагмента виконується N прогонів генерації з ненульовою температурою, що дозволяє отримати кілька варіантів коду. Варіативність результатів генерації σ_{SS} відображає невизначеність моделі.

Третій етап полягає в інтеграції отриманих показників із метрикою якості IAQS [3] у розширену метрику IAQS+ за формулою: $IAQS^+ = \alpha \cdot IAQS + \beta \cdot CS + \gamma \cdot (1 - \sigma_{SS})$, де σ_{SS} – нормована варіативність результатів стохастичного семпльовання; α , β , γ – вагові коефіцієнти

($\alpha + \beta + \gamma = 1$), що визначаються експертно або методом аналізу ієрархій.

За результатами оцінювання IAQS+ фрагменти мігрованого користувацького коду автоматично розподіляються на три категорії: «прийнято» ($IAQS+ > 0.75$), «потребує перевірки» ($0.5 \leq IAQS+ \leq 0.75$) та «відхилено» ($IAQS+ < 0.5$). Остаточне калібрування порогів планується за результатами експерименту. Така тривірнева класифікація суттєво скорочує обсяг ручної верифікації порівняно з повним переглядом усього мігрованого користувацького коду фахівцем.

Експериментальну перевірку методу заплановано на вибірці користувацького АВАР-коду з використанням моделі Qwen 2.5 Coder 14B. Очікується суттєве скорочення частки хибно прийнятих фрагментів мігрованого коду порівняно з використанням лише IAQS.

Розроблюваний метод оцінювання невизначеності LLM забезпечить формалізований контроль генерації мігрованого користувацького АВАР-коду. У подальшому планується його інтеграція в процес інтелектуального реінжинірингу КС.

Список використаних джерел:

1. Поздняков О. А. Міграція користувацького коду на нові версії складних комп'ютерних систем з використанням методів і моделей інтелектуального реінжинірингу / О. А. Поздняков, А. В. Пархоменко // Наукові праці ДонНТУ. Серія «Інформатика, кібернетика та обчислювальна техніка». — 2025. — № 2(41). — С. 86–98.

2. Поздняков О. А. Дослідження та вибір великих мовних моделей для автоматизації міграції АВАР-коду / О. А. Поздняков, А. В. Пархоменко // Управління розвитком складних систем. — 2025. — № 63. — С. 191–200.

3. Pozdnyakov O. A. Evaluation and quality assurance of migrated АВАР code using an integral metric and generative artificial intelligence models / O. A. Pozdnyakov, A. V. Parkhomenko // Computer Science. Software Engineering. — 2026. — № 1. — P. 80–89.

4. Detecting hallucinations in large language models using semantic entropy / [S. Farquhar, J. Kossen, L. Kuhn, Y. Gal] // Nature. — 2024. — Vol. 630. — P. 625–630.

5. Look before you leap: An exploratory study of uncertainty analysis for large language models / [Y. Huang, J. Song, Z. Wang et al.] // IEEE Transactions on Software Engineering. — 2025. — Vol. 51, № 2. — P. 413–429.