

# ЗНАННЯ-ОРІЄНТОВАНА ІЄРАРХІЧНА МУЛЬТИАГЕНТНА ІНТЕЛЕКТУАЛЬНА СИСТЕМА СЦЕНАРНОГО ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ НА ОСНОВІ LLM

<sup>1</sup>Вінницький національний технічний університет

*Підвищення точності, надійності та узгодженості сценарного прогнозування часових рядів з високою невизначеністю є однією з ключових проблем сучасних систем підтримки прийняття рішень. Особливу складність такі задачі становлять за наявності пропусків даних, лагових ефектів, змінної сезонності, структурних зламів і необхідності врахування зовнішніх контекстних знань, які часто представлені у текстовій формі. Традиційні методи прогнозування, орієнтовані на оптимізацію окремих моделей, зазвичай не забезпечують узгодженості прогнозів і коректного оцінювання невизначеності в сценарному аналізі.*

*У роботі запропоновано знання-орієнтовану ієрархічну мультиагентну інтелектуальну систему (IMAIC) сценарного прогнозування часових рядів, у якій великі мовні моделі (LLM) використовуються як спеціалізовані інтелектуальні агенти з чітко розподіленими ролями, які не перетинаються. На основі структурно-функціональної декомпозиції та системного підходу виділено три основні класи (рівні ієрархії) інтелектуальних агентів: агенти підготовки даних та формування ознак, агенти побудови та налаштування прогнозних моделей і агенти кооперації та оцінювання невизначеності прогнозів моделей. Запропоновано формалізацію архітектур IMAIC та інтегральний багатокритеріальний показник якості, що враховує ефективність підготовки даних, адекватність прогнозування динаміки та каліброваність зон невизначеності.*

*Сформульовано гіпотези щодо переваг ролівої декомпозиції агентів і доведено теорему про Парето-оптимальність ієрархічної IMAIC зі спеціалізованими класами агентів за середньозваженим інтегральним критерієм якості з додатними вагами. На основі цієї формалізації розроблено відповідний сценарно-орієнтований агент-бенчмарк IMAS-SCOPE (Intelligent Multi-Agent Systems — Scenario Consistency & Optimal Prediction Evaluation), призначений для порівняння альтернативних архітектур мультиагентних систем прогнозування.*

*Наведено приклад експериментальної реалізації IMAIC у середовищі Kaggle для сценаріїв на основі реальних даних першої хвилі захворюваності на COVID-19 в Україні. Показано, що для прогнозу різкого спаду після стрімкого зростання приросту кількості нових хворих повноцінна сценарна архітектура A3 забезпечує кращу у 2—3 рази точність прогнозування за метрикою WAPE, усередненою за 2 тижні, кращу на 31 % узгодженість оцінювання і каліброваність невизначеності та краще на 36 % значення інтегрального критерію корисності порівняно з базовою архітектурою A1 за сценарієм прогнозування різкого спаду одразу після стрімкого зростання. Отже, на прикладі доведено підвищення точності (зменшення WAPE), надійності (більша каліброваність зон невизначеності) та узгодженості прогнозів (за даними різних часових інтервалів), що в цілому забезпечує вибір системи з вищою якістю прогнозування і може бути також інтерпретовано і як вибір системи з вищою достовірністю прогнозів.*

*Отримані результати підтверджують доцільність використання LLM-керованих мультиагентних архітектур і спеціалізованих бенчмарків для задач сценарного прогнозування в умовах високої невизначеності.*

**Ключові слова:** мультиагентні системи, штучний інтелект, сценарне прогнозування, часові ряди, великі мовні моделі, невизначеність, ансамблеве моделювання, коронавірус.

## Вступ

Прийняття рішень у складних системах у сферах охорони здоров'я, енергетики, економіки, екології тощо, зазвичай, ґрунтується на аналізі і прогнозуванні часових рядів та сценарному прогно-

зуванні. При цьому важливо забезпечувати точність, надійність та узгодженість цих прогнозів, не дивлячись на те, що часові ряди мають суттєву різного роду невизначеність, неповноту, пропуски даних та їх суперечливість. Наприклад, нещодавній практичний досвід управління епідемічними процесами під час пандемії коронавірусу у світі та, зокрема в Україні, виявив багато проблем. Зокрема й щодо невизначеності даних, причому чим більше випадків захворювання людей і ціннішими є вимоги до точних прогнозів, тим стає важче збирати достовірну інформацію, оскільки хворіють ті люди, які мали збирати, обробляти та передавати цю інформацію, а це, також призводить до неточних прогнозів, запізненого чи передчасного запровадження обмежувальних заходів, неефективного планування медичних ресурсів, некоректного оцінювання ризиків поширення інфекції тощо [1]—[5]. Пандемія COVID-19 стала показовим прикладом задачі, в якій прогнозування здійснюється в умовах неповних, зашумлених та структурно нестабільних даних.

Упродовж останніх років запропоновано велику кількість підходів до аналізу і прогнозування часових рядів — від класичних ARIMA-моделей до ансамблевих методів машинного навчання, епідеміологічних моделей, баєсівських підходів і Prophet-моделей. Проте, більшість із них орієнтована на оптимізацію окремої моделі та демонструє знижену ефективність за умов високої невизначеності, наявності структурних зламів, складної сезонності, затримок даних. Додатковим ускладненням є те, що в реальних задачах прогнозування разом із числовими часовими рядами активно використовуються текстові аналітичні звіти, експертні оцінки та пояснювальні матеріали, які містять причинно-наслідкові, контекстні та логічні залежності, що не представлені у числовій формі.

З появою великих мовних моделей (LLM) сформувався новий напрям досліджень, пов'язаний із поєднанням текстових знань і числових моделей для аналізу та прогнозування часових рядів. Наприклад, активно розвиваються напрями Time-LLM, LLM-орієнтованого імпутигу та семантичного аналізу часових даних у задачах виявлення структурних властивостей даних, вибору релевантних ознак і параметрів моделей [1], [2], [4], [6]—[9]. Водночас, переважна більшість таких моделей, методів та технологій зосереджена на удосконаленні окремих складових прогнозів і не розглядає узгоджену роботу кількох спеціалізованих агентів у межах складної ієрархічної системи.

У сфері створення та розвитку мультиагентних систем також існують підходи до підтримки прийняття рішень в умовах невизначеності [10], [11], проте вони, зазвичай, не орієнтовані на сценарне прогнозування часових рядів, не достатньо інтегрують текстові знання з використанням інтелектуальних агентів з LLM. А головне, відсутні стандартизовані агент-бенчмарки, які дозволяли б коректно порівнювати різні мультиагентні архітектури саме з позицій сценарності, узгодженості прогнозів і каліброваності оцінювання невизначеності.

У зв'язку з цим, актуальною є задача розроблення знання-орієнтованої ієрархічної мультиагентної інтелектуальної системи сценарного прогнозування, в якій використовуються інтелектуальні агенти з LLM як спеціалізовані агенти для видобування знань і ознак, аналізу динаміки та налаштування прогнозних моделей, а також для кооперації результатів і формування узгоджених сценаріїв з коректними зонами невизначеності.

*Метою статті* є підвищення точності, надійності та узгодженості сценарного прогнозування часових рядів з високою невизначеністю шляхом розроблення знання-орієнтованої ієрархічної мультиагентної інтелектуальної системи оптимальної архітектури, для обґрунтування вибору якої також варто створити відповідний сценарно-орієнтований агент-бенчмарк.

### Формалізація постановки задачі

Типова ієрархічна мультиагентна інтелектуальна система (ІМАІС), призначена для сценарного прогнозування складних динамічних процесів в умовах невизначеності, — це складна система з багатьма рівнями ієрархії інтелектуальних агентів, які використовують великі мовні моделі чи інші моделі та технології. Організація роботи такої системи, а тим більше, доведення її оптимальності за якимись критеріями є дуже складною задачею, яку навряд чи можна розв'язати в загальному вигляді. До того, більшість великих мовних моделей є універсальними дуже потужними інструментами для розв'язання широкого класу задач, але вони є, скоріше, «чорними скриньками» в модельному плані, оскільки їх структура складна і не завжди відома (наприклад, структура та усі параметри OpenAI ChatGPT 5.1 не є публічно доступними). А тому пропонуємо внести суттєві спрощення у постановку задачі:

1) розглядати тільки такі ІМАІС, в яких кожен з агентів є спеціалізованим під чітко визначену роль, причому ролі агентів не перетинаються;

2) на кожному рівні ієрархії знаходиться підсистема агентів певної ролі, наприклад, відповідно до етапів системного аналізу або машинного навчання та інтелектуального аналізу даних: видобування даних, інженерія ознак, побудова моделі, аналіз моделі, підтримка прийняття рішень;

3) на кожному рівні варто забезпечити максимум ефективності оброблення, що визначається певним комплексним кількісним критерієм.

Основна ідея розв'язання поставленої задачі полягає у функціонально-структурній декомпозиції процесу сценарного прогнозування на послідовні, логічно завершені етапи, кожному з яких відповідає окремий клас агентів. У базовому варіанті ІМАІС пропонуємо виділяти три вищезгадані основні класи (рівні) агентів:

– агенти першого класу ( $\Omega_1$ ) — здійснюють видобування, очищення, узгодження даних та формування ознак, метою їх роботи є побудова цілісного датасету  $D$  на інтервалі часу  $t = [0, T]$ , придатного для подальшого аналізу

$$D = \{(x_t, y_t)\}_{t=0}^T, \quad (1)$$

де  $x_t$  — вектор ознак у момент часу  $t$ ;  $y_t$  — значення даних спостережень, яке відповідає цим ознакам у момент часу  $t$ ;

– агенти другого класу ( $\Omega_2$ ) — будують та оптимізують окремі інтелектуальні моделі прогнозування на основі датасетів, сформованих агентами першого класу, здійснюють за ними прогнози; їх метою є формування  $K$  альтернативних сценаріїв  $S$  розвитку процесу з різним ступенем прогнозної невизначеності та горизонтом прогнозування  $H$

$$S = \left\{ s^{(k)} = \left( \hat{y}_{t+1}^{(k)}, \dots, \hat{y}_{t+H}^{(k)} \right) \right\}_{k=1}^K; \quad (2)$$

– агенти третього класу ( $\Omega_3$ ) — забезпечують кооперацію результатів роботи агентів другого класу (етапу аналізу), включно з ансамблюванням прогнозів, оцінюванням ступеня невизначеності; метою є формування узгоджених сценарних рішень та, на їх основі — рекомендацій підтримки прийняття рішень (хоча це можуть робити й агенти наступного класу або рівня).

Теоретично кількість класів агентів може бути розширена, проте у межах цієї роботи аналіз обмежується зазначеними трьома класами як мінімально достатніми для реалізації повного циклу сценарного прогнозування.

Отже, архітектуру усіх можливих ІМАІС можна формалізувати як множину послідовно з'єднаних множин агентів

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3, \quad (3)$$

а архітектуру конкретної ІМАІС  $A$  можна формалізувати у вигляді

$$A = A_1 \cup A_2 \cup A_3, \quad A_1 \in \Omega_1, \quad A_2 \in \Omega_2, \quad A_3 \in \Omega_3. \quad (4)$$

Ефективність функціонування ІМАІС пропонується оцінювати за допомогою середньозваженого інтегрального критерію якості  $J$ , який набуває значень в діапазоні  $[0, 1]$

$$J = \sum_{i=1}^N w_i U_i; \quad \sum_{i=1}^N w_i = 1, \quad \forall w_i > 0, \quad (5)$$

де  $U_i$  — комплексний критерій зі значеннями в діапазоні  $[0, 1]$  з вагою  $w_i$ , який відображає якість оброблення даних усіма агентами на  $i$ -му рівні архітектури агента  $A$ ;  $N$  — кількість рівнів ІМАІС, в цьому випадку обмежуємось  $N = 3$ .

Формально, запропонований інтегральний критерій якості прогнозування може бути також розглянутий як критерій оцінювання ступеня достовірності прогнозів системи, тобто як засіб кількісного оцінювання рівня обґрунтованої довіри до прогнозних результатів.

Задачею оптимізації є досягнення максимальної ефективності ІМАІС, отже потрібно доставити максимум критерію (5), забезпечуючи максимум кожній його складовій.

Критерій  $U_1$  враховує результативність процесів видобування інформації, формування ознак, передоброблення і підготовки датасету. Доволі детально операції цього етапу описані у роботах [12], [13]. Він може оцінюватися за допомогою різних метрик, зокрема похибок відновлення даних, збереження статистичних властивостей або стабільності побудованих ознак. Наприклад, це може бути похибка (MSE, RMSE тощо [12]) імпутингу пропущених даних, яка обчислюється на валідаційному

датові і нормується у певний спосіб, але, щоб мінімізація похибки забезпечувала максимізацію критерію, можна використати, наприклад таку функцію:

$$U_1 = \frac{1}{1 + e_1}, \quad (6)$$

де  $e_1$  — нормоване значення похибки імпутації валідаційних даних.

Критерій  $U_2$  відображає адекватність побудови моделей та адекватність вибраної оптимальної серед них, чи — адекватність декількох найкращих, які можна поєднати в ансамбль. Агенти цього рівня, у разі роботи з часовими рядами, повинні якісно відтворювати тренд і сезонні складові, з урахуванням можливих аномальних значень, моделювати волатильність та локальні коливання ряду, виявляти та адаптуватись до структурних змін і режимних переходів процесів. Для сценарного прогнозування, в ширшому сенсі, важливою є здатність системи своєчасно (іноді — й завчасно) реагувати на зміну умов та коректно моделювати альтернативні траєкторії розвитку процесу. Зазвичай, цей критерій враховує те, як точно найкраще ідентифікована на цьому рівні ієрархії модель прогнозує дані валідаційного датасету. Це, зазвичай, середньозважена сума різного роду метрик (MAE, MSE, RMSE, WAPE, SMAPE тощо [12]), кожна з яких враховується у подібний до (6) спосіб. Або це може бути й метрика  $R_2$  (для регресійної задачі) чи  $F_1$  (для класифікаційної задачі), в яких що більше значення, то краще, а максимум один, це тому такі метрики враховують просто з коефіцієнтом. Можуть бути й специфічніші метрики, наприклад волатильність приростів (помилка за першими різницями ряду для аналізу точності відтворення динаміки ряду) та інші.

Критерій  $U_3$  враховує як зменшення зони невизначеності (чим прогноз впевненіший, тим краще), так і каліброваність прогнозу — наскільки зона невизначеності покриває («coverage») істинні значення валідаційного датасету. У цій метриці важливо так підібрати функції штрафу, щоб критерій не мав параболічну форму, тобто, щоб на певному допустимому інтервалі визначення  $U_3$  його зростання завжди означало покращення ефективності, а зменшення — завжди погіршення. Цей критерій відображає узгодженість прогнозів системи. Важливо зазначити, що під узгодженістю сценарного прогнозування в цій роботі розуміється властивість прогнозної системи зберігати близькість прогнозних траєкторій та відповідних зон невизначеності прогнозів за допустимих варіацій вхідних даних, параметрів моделей та сценарних припущень. Вища узгодженість прогнозування зменшує ризик неконтрольованого розбігу результатів під час програвання сценаріїв та ухвалені рішення стають ціннішими з використанням таких прогнозів.

Зазвичай найціннішою для функції корисності (5) є ефективна робота агентів другого класу, а найменш цінною — першого. Хоча, звичайно, це залежить і від кількості і характеру проблемних даних та специфіки задачі: гарний прогноз можна отримати тільки з якісних даних. Якщо дані неякісні, то це відіб'ється на усіх наступних етапах їх оброблення. Тому ваги інтегрального критерію потрібно вибрати теж оптимально.

Доповнимо введені спрощення та обмеження ще двома гіпотезами і лемою.

**Гіпотеза 1.** ІМАІС з чітким розподілом ролей, що не перетинаються, між її агентами підвищує значення критерію  $U_2$  у задачах сценарного прогнозування складних динамічних процесів.

**Гіпотеза 2.** Використання механізмів кооперації та оцінювання невизначеності викликає зростання критерію  $U_3$  порівняно з ізольованим використанням окремих моделей.

**Лема 1.** Поліпшення результативності агентів будь-якого класу без погіршення результативності інших класів агентів приводить до зростання відповідної складової інтегрального критерію  $J$ .

Лема 1 впливає із характеру та опуклості критерію (5).

А тоді має місце теорема.

**Теорема.** Ієрархічна ІМАІС зі спеціалізованими класами агентів та неперетинними ролями є Парето-оптимальною за інтегральним критерієм  $J$ , тобто не існує іншої архітектури, яка забезпечувала б одночасне покращення всіх складових  $U_1, U_2, U_3$  без погіршення хоча б однієї з них.

**Доведення.** Легко показати, що функціональна декомпозиція задачі на незалежні етапи з окремими класами агентів мінімізує конфлікти між цілями оптимізації різних складових критерію. Відтак, покращення якості роботи одного класу агентів не потребує деградації результатів агентів інших класів, що й забезпечує Парето-оптимальність архітектури для задач сценарного прогнозування.

## Можливі Парето-оптимальні архітектури ІМАІС для сценарного прогнозування

Наведемо декілька архітектур (4), які задовольняють усім вищенаведеним вимогам.

1. *Архітектура А1* — LLM-керована локальна мультиагентна система.

LLM-агенти класу  $U_1$  розв'язують такі задачі: семантичний аналіз доступних джерел даних, видобування необроблених даних  $D_{raw}$ , автоматизований пошук потенційно релевантних ознак, формування обмеженого датасету  $D$  на основі історичних значень

$$D = L_1(D_{raw}, P_1),$$

де  $L_1(\cdot)$  — LLM-агент 1-го рівня (класу);  $P_1$  — промпти 1-го рівня.

LLM-агенти класу  $U_2$  розв'язують такі задачі: базовий аналіз динаміки рядів, налаштування параметрів однієї домінуючої прогнозної моделі.

Прогнозна модель

$$\hat{y}_{t+h} = f_{\theta}(x_t), \quad \theta = L_2(D, P_2),$$

де  $L_2$  — LLM-агент 2-го рівня (класу), що здійснює тюнінг параметрів  $\theta$ .

LLM-агенти класу  $U_3$  розв'язують такі задачі: формування простої зони невизначеності  $U_{t+h}$  на основі залишкової дисперсії  $\sigma^2$

$$\sigma^2 = \text{var}(y_t - \hat{y}_t); \quad U_{t+h} = \hat{y}_{t+h} \pm z_{\alpha} \sigma.$$

2. *Архітектура А2* — LLM-агентна система лише зі структурним аналізом динаміки.

LLM-агенти  $U_1$  — виконують автоматизований опис часової структури, виділяють трендові  $\tau_t$  та сезонні  $\varphi_t$  компоненти

$$D = L_{11}(D_{raw}, P_{11}); \quad (\tau_t, \varphi_t) = L_{12}(D_{raw}, P_{12}).$$

LLM-агенти  $U_2$  — здійснюють пошук аномалій, оцінюють параметри сезонності, адаптують структуру прогнозної моделі.

LLM-агенти  $U_3$  — формують часово-залежну оцінку волатильності.

3. *Архітектура А3* — повноцінна LLM-керована сценарна ІМАІС.

LLM-агенти першого класу інтегрують числові, текстові та зовнішні індикатори, формують множину альтернативних представлень даних у різного роду датасетах

$$\{D_m\}_{m=1}^M = L_{m1}(D_{raw}, P_{m1}).$$

LLM-агенти другого класу виявляють аномалії та точки зміни режимів, підбирають параметри сезонності, керують тюнінгом моделей для різних сценаріїв.

$$S = \{s^{(k)}\}; \quad s^{(k)} = M_k(D_m, \theta_k); \quad \theta_k = L_{m2}(D_m, P_{m2}).$$

Клас  $U_3$  — LLM-агенти кооперації прогнозів та оцінювання їх каліброваної невизначеності.

LLM-агенти третього класу аналізують узгодженість сценаріїв, формують правила ансамблювання, оцінюють калібровану невизначеність.

Ансамблевий прогноз значення  $\bar{y}_{t+h}$

$$\bar{y}_{t+h} = \sum_{k=1}^K w_k \hat{y}_{t+h}^{(k)}, \quad w_k = L_{m3}(S, P_{m3}).$$

## Вимоги до бенчмарку для вибору оптимальної архітектури ІМАІС

На основі умов, лем та теореми, сформульованих у попередніх розділах, бенчмарк для вибору оптимальної архітектури ІМАІС повинен мати відповідну структуру для тестування кожного виду множин агентів у (3), зокрема він має задовольняти такі вимоги:

1. *Рольова декомпозиція тестів.*

Бенчмарк має бути побудований таким чином, щоб результати роботи кожного класу LLM-агентів могли бути оцінені незалежно. Це означає, що:

– тестові завдання повинні містити елементи, чутливі до якості видобування даних і формування ознак ( $U_1$ );

– окремі сценарії мають перевіряти здатність моделей до аналізу динаміки, сезонності, аномалій

та адаптації ( $U_2$ );

– повинні бути передбачені умови для оцінювання кооперації моделей і каліброваності невизначеності ( $U_3$ ).

## 2. Сценарна варіативність.

З огляду на те, що архітектури ІМАІС орієнтовані на сценарне прогнозування, бенчмарк має містити множину сценаріїв, які відрізняються за:

- характером динаміки;
- рівнем волатильності;
- наявністю аномалій та структурних зламів;
- ступенем невизначеності.

Ця вимога забезпечує можливість перевірки гіпотези про перевагу сценарно-орієнтованих архітектур і є необхідною умовою для виявлення Парето-оптимальності.

## 3. Порівнюваність та нормування метрик.

Метрики, що використовуються для оцінювання  $U_1$ ,  $U_2$  та  $U_3$ , мають бути приведені до порівнюваної шкали, наприклад до інтервалу  $[0, 1]$ . Це дозволяє:

- формувати векторні оцінки якості;
- аналізувати Парето-фронт альтернативних архітектур;
- обчислювати інтегральний критерій корисності  $J$ .

Без нормування та уніфікації шкал порівняння архітектур втрачає коректність.

## 4. Оцінювання невизначеності як окремого об'єкта тестування.

Згідно з теоремою про Парето-оптимальність, архітектура не може вважатися оптимальною, якщо покращення точності прогнозу досягається ціною некоректного оцінювання невизначеності.

Тому бенчмарк має містити спеціальні тести, спрямовані на:

- оцінювання каліброваності прогнозних інтервалів;
- аналіз стабільності зон невизначеності зі зміною сценаріїв;
- перевірку узгодженості ансамблю прогнозних моделей.

## 5. Повторюваність та стійкість результатів.

Для перевірки стійкості архітектури до випадкових факторів бенчмарк має передбачати:

- багаторазове програвання сценаріїв;
- варіювання початкових умов;
- оцінювання розсіювання метрик якості.

Це дозволяє уникнути вибору архітектури, оптимальної лише для окремого набору умов.

З огляду на універсальність запропонованої формалізації, для побудови тестів бенчмарку можуть використовуватися як синтетичні, так і реальні дані.

Синтетичні часові ряди дозволяють:

- керувати задавати тренди, сезонність і волатильність;
- моделювати сценарії зі штучно введеними аномаліями та структурними зламами;
- перевіряти поведінку ІМАІС у граничних та рідкісних випадках.

Такі дані є особливо корисними для валідації окремих складових критеріїв  $U_1$ — $U_3$ .

Реальні дані дозволяють:

- перевірити можливість та ефективність практичної реалізації архітектур;
- оцінити роботу LLM-агентів в умовах неповної, зашумленої та неоднорідної інформації;
- порівняти ефективність роботи агентів з еталонними (гарними і не дуже) аналогами;
- підтвердити узгодженість теоретичних положень з практичними результатами.

Таким чином, бенчмарк для вибору оптимальної архітектури ІМАІС має бути сценарно-орієнтованим, рольово-декомпованим та багатокритеріальним. Лише з виконанням цих вимог можливе коректне виявлення Парето-оптимальних архітектур і обґрунтований вибір найефективнішої ІМАІС для задач сценарного прогнозування.

## Приклад застосування методу

Продемонструємо приклад застосування запропонованої технології на прикладі моделювання сценаріїв зміни щоденних приростів кількості нових підтверджених хворих на коронавірус (COVID-19) під час першої великої хвилі в Україні у другій половині 2020 р. Як відомо, при Президії НАН України працювала Робоча група, в якій одна команда вчених здійснювала прогнозування таких даних шляхом удосконалення їх моделі SEIR-U [14], а друга — моделі на базі Prophet (другою командою

керував один зі співавторів цієї статті — Мокін В. Б.) [3]. Усі результати кожні 1—2 тижні узагальнювались у звіти, які розсилались в РНБО України та інші урядові інституції, а також викладались у публічний доступ на сайті Президії НАН України [15]. На початку грудня 2020 року вперше спостерігалась відносно велика хвиля кількості хворих і тому мав місце невдалий прогноз обома моделями [16], коли сценарій стрімкого зростання досяг піку 28.11.2020 р. і різко пішов на спад. Тоді ті моделі дали похибку більше 30 % на 2 тижні, які мали місце одразу після проходження піку хвилі [5]. Цей сценарій («різкий спад після стрімкого зростання») тоді ніхто вдало не спрогнозував, хоча модель SEIR-U, яка враховувала зміни репродуктивного числа [14], спрогнозувала уповільнення зростання і, навіть деяке зменшення, але недостатньо для точного прогнозу [5].

Для ілюстрації запропонованого підходу реалізовано та порівняно дві альтернативні архітектури ієрархічної мультиагентної інтелектуальної системи сценарного прогнозування — A1 та A3, які відповідають різним рівням сценарної виразності та кооперації агентів.

Як зазначено вище, архітектура A1 (базова LLM-керована мультиагентна система) реалізує спрощений варіант ІМАІС, у якому всі три класи LLM-агентів присутні, проте їх функціональність сильно обмежена.

LLM-агенти класу  $U_1$  виконують відбір обмеженого набору релевантних ознак та формування єдиного датасету для прогнозування  $D$ .

LLM-агенти класу  $U_2$  здійснюють налаштування параметрів однієї прогнозувальної моделі та виконують прогнозування без явного формування альтернативних сценаріїв (відповідно до критерію  $U_2$ ).

LLM-агенти класу  $U_3$  формують просту зону невизначеності на основі залишкової дисперсії прогнозувальної моделі.

Таким чином, архітектура A1 орієнтована переважно на мінімізацію середньої похибки прогнозу, але має обмежені можливості сценарного аналізу та кооперації моделей.

Наведена вище архітектура A3 (повноцінна LLM-керована сценарна ІМАІС) реалізує запропонований у роботі підхід до знання-орієнтованого сценарного прогнозування.

LLM-агенти класу  $U_1$  формують множину альтернативних представлень даних, зокрема шляхом вибору різних вікон історії, наборів ознак та зовнішніх індикаторів.

LLM-агенти класу  $U_2$  виконують аналіз динаміки часових рядів, виявлення аномалій, параметрів сезонності та налаштування параметрів прогнозувальних моделей для кожного сценарію окремо.

LLM-агенти класу  $U_3$  забезпечують кооперацію результатів прогнозування, формування ансамблевого сценарію та калібрування зон невизначеності.

На відміну від A1, архітектура A3 орієнтована не лише на точність прогнозу, а й на адекватне відтворення динаміки, волатильності та змін сценаріїв, що безпосередньо відповідає постановці задачі сценарного прогнозування.

Для вибору оптимальної архітектури на прикладі реальних даних охарактеризованої вище задачі прогнозування поширення коронавірусу в Україні розроблено сценарно-орієнтований бенчмарк ІМАС-SCOPE, що відповідає вимогам, сформульованим у попередньому розділі.

Бенчмарк включає:

- множину ковзних сценаріїв прогнозування з фіксованим горизонтом;
- варіативність динаміки часових рядів (тренд, сезонність, волатильність);
- перевірку точності прогнозування та якості оцінювання невизначеності.

Для кожної архітектури обчислювались складові векторного критерію якості:

$U_1$  — метрики стабільності та узгодженості підготовлених даних і ознак;

$U_2$  — метрики точності прогнозування та відтворення динаміки, серед яких основною була метрика, яка, певною мірою для позитивних цілих значень (кількість людей) є аналогом відносною похибки прогнозу (WAPE, усереднена за 2 тижні, помножена на 100, у %) та показники динамічної відповідності;

$U_3$  — метрики каліброваності зон невизначеності, зокрема частка покриття істинних значень та відхилення від заданого рівня покриття.

На основі нормованих значень обчислювався інтегральний критерій корисності  $J$ , який максимізується.

Інструменти оцінювання сформовано відповідно до структури бенчмарку та формалізації задачі й містили:

- автоматизоване з використанням LLM формування сценарних вікон прогнозування;
- обчислення метрик якості для кожного сценарію;
- агрегацію результатів за різними сценаріями для кожного виду архітектури окремо;

- формування табличних та графічних представлень результатів.
- Оцінювання проводилось окремо для кожного сценарію, що дозволило:
  - аналізувати поведінку архітектур у різних динамічних умовах;
  - уникнути упередженості, пов’язаної з одиничним експериментом;
  - коректно порівнювати архітектури за векторним критерієм якості.

На рис. 1, 2 наведено приклади, які ілюструють результати роботи ІМАІС з архітектурами А1 та А3.

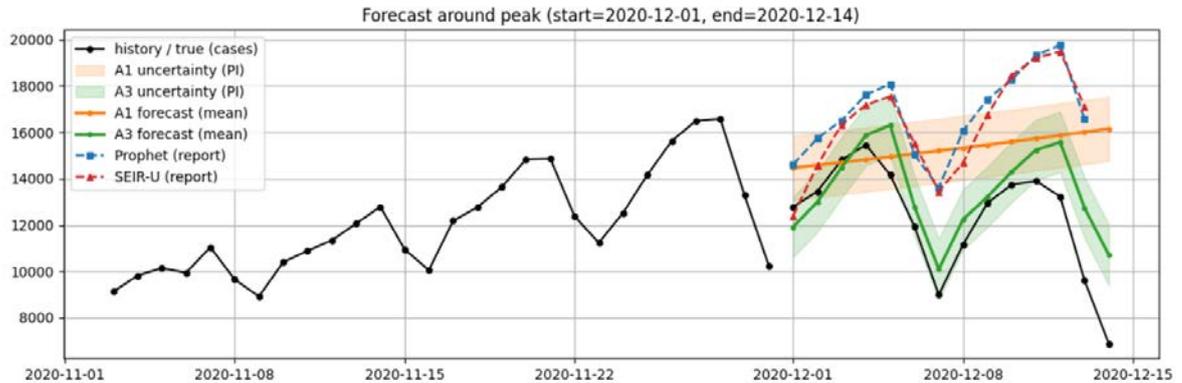


Рис. 1. Сценарний прогноз часових рядів з порівнянням архітектур А1 та А3 на період після піку, а також базових моделей SEIR-U та моделі на базі Prophet; показано історичні дані, прогнозні траєкторії та зони невизначеності

arch	U2_wape	U1	U2	U3	J
A1	24.517	0.904	0.558	0.660	0.623
A3	10.821	0.985	0.817	0.864	0.848

Рис. 2. Значення метрики WAPE для прогнозів, усереднене за 2 тижні, для моделей агентів 2-го класу (рівня), усіх  $U_1$ — $U_3$  складових критерію (5) та інтегрального критерію  $J$  в цілому для кожної з архітектур А1 і А3 на окремому сценарному інтервалі прогнозування «різкий спад після стрімкого зростання» (після проходження піку)

Аналіз рис. 1 і 2 підтверджує, що архітектура А3 забезпечує кращу точність прогнозування за середньою за 2 тижні метрикою WAPE, краще відтворення динаміки, узгодженість оцінювання та каліброваність невизначеності. Зокрема, для А3 у порівнянні з А1 має місце зменшення «U2\_wape» у 2,3 разів (та у 3 рази краще, ніж у моделі SEIR-U та моделі на базі Prophet [5]), більші значення критеріїв  $U_1$ — $U_3$  у 1,1—1,5 рази та більше значення інтегрального критерію  $J$  на 36 %. Найбільше на зростання  $J$  вплинуло зростання критерію  $U_2$  на 46,4 %, який був урахований з вагою  $w_2 = 0,6$ .

Таким чином, реалізований приклад підтверджує теоретичні положення щодо Парето-оптимальності сценарно-орієнтованої LLM-керованої ІМАІС та доцільності використання спеціалізованого агент-бенчмарку для вибору оптимальної архітектури. Це був лише один приклад формування та застосування еталонних тестів агент-бенчмарку. Для об’єктивного висновку про Парето-оптимальність архітектур А1 та А3 потрібні десятки і сотні тисяч різноманітних подібних прикладів.

## Висновки

У роботі розроблено знання-орієнтовану ієрархічну мультиагентну інтелектуальну систему (ІМАІС) сценарного прогнозування часових рядів з високою невизначеністю, в якій великі мовні моделі використовуються як спеціалізовані агенти з чітко визначеними ролями, що не перетинаються. Запропоновано формалізацію архітектур ІМАІС, що базується на функціонально-структурній декомпозиції процесу прогнозування на етап підготовки даних, етап побудови моделей і прогнозів та етап ансамблювання прогнозів і сценарного аналізу.

Наукова новизна роботи полягає у формулюванні інтегрального багатокритеріального показника якості ІМАІС та доведенні теореми про Парето-оптимальність ієрархічної архітектури зі спеціалізованими класами агентів. Обґрунтовано та розроблено структуру і вимоги до нового сценарно-орієнтованого агент-бенчмарка IMAS-SCOPE, який дозволяє коректно порівнювати альтернативні мультиагентні архітектури з урахуванням точності прогнозування, відтворення динаміки та каліброваності невизначеності.

Експериментальна апробація на реальних даних COVID-19 в Україні підтвердила, що повноцінна

LLM-керована сценарна архітектура А3 забезпечує краще відтворення динаміки (вищу у 2—3 рази точність прогнозування за метрикою WAPE та більше на 46,4 % значення ефективності агентів другого рівня), кращу узгодженість оцінювання і каліброваність невизначеності (більше на 30,9 % значення ефективності агентів третього рівня) та краще на 36,1 % значення інтегрального критерію ніж базова архітектура А1. Отримані результати узгоджуються з теоретичними положеннями роботи та підтверджують доцільність використання ІМАІС для задач сценарного прогнозування. Отже, у межах сценарного бенчмарку, що передбачає багаторазове прогнозування на різних часових інтервалах з варіаціями вхідних умов, показано підвищення точності (зменшення значення WAPE), надійності (більша каліброваність зон невизначеності) та узгодженості прогнозів (за даними різних часових інтервалів), що кількісно підтверджує досягнення мети роботи. До того ж, з огляду на комплексний характер використаного інтегрального критерію, вибір системи з вищою якістю прогнозування, може бути також інтерпретовано і як вибір системи з вищою достовірністю прогнозів.

Подальші дослідження доцільно спрямувати на розширення кількості класів агентів, формалізацію механізмів взаємодії між LLM-агентами та розроблення універсальних сценарних бенчмарків для інших прикладних галузей.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] F. Petropoulos, et al., “COVID-19: Forecasting confirmed cases and deaths with a simple time series model,” *International Journal of Forecasting*, vol. 38, no. 2, pp. 439-452, Apr.–Jun. 2022. <https://doi.org/10.1016/j.ijforecast.2020.11.010>.
- [2] A. Ghosh, et al., “Mathematical modelling for decision making of lockdown during COVID-19,” *Applied Intelligence*, vol. 52, pp. 699-715, Jan. 2022. <https://doi.org/10.1007/s10489-021-02463-7>.
- [3] В. Б. Мокін, А. В. Лосенко, і А. Р. Яцолт, «Інформаційна технологія аналізу та прогнозування кількості нових випадків захворювань на коронавірус SARS-CoV-2 в Україні на основі моделі Prophet,» *Вісник Вінницького політехнічного інституту*, № 5, с. 71-83, Листоп. 2020. <https://doi.org/10.31649/1997-9266-2020-152-5-71-83>.
- [4] Z. Qu, et al., “Forecasting New COVID-19 Cases and Deaths Based on an Intelligent Point and Interval System Coupled With Environmental Variables,” *Frontiers in Ecology and Evolution*, vol. 10, 2022. <https://doi.org/10.3389/fevo.2022.875000>.
- [5] А. В. Лосенко, «Інформаційна технологія прогнозування часового ряду кількості хворих на коронавірус на основі моделі Facebook Prophet,» *Вісник Вінницького політехнічного інституту*, № 5, с. 50-59, Жовт. 2023. <https://doi.org/10.31649/1997-9266-2023-170-5-50-59>.
- [6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, 2015, 712 p. ISBN: 978-111867502.
- [7] Y. Jiang, et al., “Empowering Time Series Analysis with Large Language Models: A Survey,” in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, 2024, pp. 8095-8103. [Electronic resource]. Available: <https://www.ijcai.org/proceedings/2024/0895.pdf>.
- [8] S. Abdullahi, et al., “Time-Series Large Language Models: A Systematic Review of State-of-the-Art,” *IEEE Access*, vol. 13, pp. 30235-30261, 2025. <https://doi.org/10.1109/ACCESS.2025.3535782>.
- [9] M. Jacobsen, and M. Tropmann-Frick, “Imputation Strategies in Time Series Based on Language Models,” *Datenbank-Spektrum*, vol. 24, pp. 197-207, Nov. 2024. <https://doi.org/10.1007/s13222-024-00488-z>.
- [10] K. Motlaghzadeh, et al., “A multi-agent decision-making framework for evaluating water and environmental resources management scenarios under climate change,” *Science of the Total Environment*, vol. 864, 2023. <https://doi.org/10.1016/j.scitotenv.2022.161060>.
- [11] L. P.-Y. Ting, et al., “Towards Hierarchical Multi-Agent Decision-Making for Uncertainty-Aware EV Charging,” *arXiv preprint arXiv:2412.18047*, Dec. 2024. [Electronic resource]. Available: <https://arxiv.org/abs/2412.18047>. <https://doi.org/10.48550/arXiv.2412.18047>.
- [12] В. Б. Мокін, і М. В. Драгований, *Наука про дані: машинне навчання та інтелектуальний аналіз даних*, електронний навч. посіб. комбінованого (локального та мережевого) використання. [Електронний ресурс]. Вінниця, Україна: ВНТУ, 2024, 258 с.
- [13] М. В. Драгований, і В. Б. Мокін, «Інтелектуальний метод з підкріпленням синтезу оптимального конвеєру операцій попереднього оброблення даних у задачах машинного навчання», *Наукові праці ВНТУ*, № 4, Черв. 2023. <https://doi.org/10.31649/2307-5392-2022-4-15-24>.
- [14] І. О. Бровченко, «Розробка математичної моделі поширення епідемії COVID-19 в Україні,» *Світгляд*, № 2 (82), с. 2-14, 2020. [Електронний ресурс]. Режим доступу: <https://files.nas.gov.ua/PublicMessages/Documents/0/2020/05/200506172747204-403.pdf>.
- [15] Президія Національної академії наук України, *Прогноз розвитку епідемії COVID-19 в Україні*. [Електронний ресурс]. Режим доступу: <https://old.nas.gov.ua/UA/Activity/covid/Pages/wg.aspx>. Дата звернення: Листоп. 2025.
- [16] Президія Національної академії наук України, *Прогноз розвитку епідемії COVID-19 в Україні на 1–14 грудня 2020 року (“Прогноз РГ-30”)*, Груд. 2020. [Електронний ресурс]. Режим доступу: <https://old.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7215>. Дата звернення: Листоп. 2025.

Рекомендована кафедрою системного аналізу та інформаційних технологій ВНТУ

Стаття надійшла до редакції 20.12.2025

**Варер Борис Юхимович** — аспірант кафедри системного аналізу та інформаційних технологій, e-mail: androbor17@gmail.com ;

**Мокін Віталій Борисович** — д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій; e-mail: vbmokin@vntu.edu.ua .

Вінницький національний технічний університет, Вінниця

**B. Yu. Varer**<sup>1</sup>

**V. B. Mokin**<sup>1</sup>

## **Knowledge-Oriented Hierarchical Multi-Agent Intelligent System for Scenario Time Series Forecasting Based on LLM**

<sup>1</sup>Vinnitsia National Technical University

*Improving the accuracy, reliability, and consistency of scenario-based forecasting of time series with high uncertainty is one of the key challenges of modern decision support systems. Such tasks are particularly complex in the presence of missing data, lag effects, variable seasonality, structural breaks, and the need to account for external contextual knowledge, which is often represented in textual form. Traditional forecasting methods focused on optimizing individual models usually fail to ensure forecast consistency and proper uncertainty estimation in scenario analysis.*

*This paper proposes a knowledge-oriented hierarchical multi-agent intelligent system (IM AIS) for scenario-based time series forecasting, in which large language models (LLMs) are used as specialized intelligent agents with clearly distributed, non-overlapping roles. Based on a structural-functional decomposition and a systems approach, three main classes (hierarchical levels) of intelligent agents are identified: data preparation and feature formation agents, forecasting model construction and tuning agents, and agents for cooperation and assessing the uncertainty of model forecasts. Formalization of IM AIS architectures and an integral multicriteria quality indicator are proposed, taking into account the effectiveness of data preparation, the adequacy of dynamic forecasting, and the calibration of uncertainty regions.*

*Hypotheses regarding the advantages of role-based agent decomposition are formulated, and a theorem on the Pareto optimality of a hierarchical IM AIS with specialized classes of agents according to the weighted average integral quality criterion is proved. Based on this formalization, corresponding scenario-oriented agent benchmark IMAS-SCOPE (Intelligent Multi-Agent Systems — Scenario Consistency & Optimal Prediction Evaluation) is developed for comparing alternative architectures of multi-agent forecasting systems.*

*An example of the experimental implementation of IM AIS in the Kaggle environment for scenarios based on real data of the first wave of COVID-19 incidence data in Ukraine is presented. It is shown that for the forecast of a sharp decline after a rapid increase in the number of new patients, the full-fledged scenario architecture A3 provides 2.3 times better forecasting accuracy by the WAPE metric averaged over 2 weeks, 31 % better estimation consistency and uncertainty calibration, and 36% better value of the integral utility criterion compared to the basic architecture A1 for the forecasting scenario of a sharp decline immediately after a rapid increase. Thus, the numerical example demonstrates improved forecasting accuracy (reduced WAPE), enhanced reliability (better calibration of uncertainty intervals), and increased forecast consistency across different time intervals, which jointly support the selection of a forecasting system with higher predictive quality and forecast credibility.*

*The obtained results confirm the feasibility of using LLM-driven multi-agent architectures and specialized benchmarks for scenario-based forecasting tasks under high uncertainty.*

**Keywords:** multi-agent systems, artificial intelligence, scenario forecasting, time series, large language models, uncertainty, ensemble modeling, coronavirus.

**Varer Borys Yu.** — Post-Graduate Student of the Chair of System Analysis and Information Technologies, e-mail: androbor17@gmail.com ;

**Mokin Vitalii B.** — Dr. Sc. (Eng.), Professor, Head of the Chair of System Analysis and Information Technologies, e-mail: vbmokin@vntu.edu.ua