Claude в AWS:

ШІ для бізнесу без ризику для даних

Як зробити бізнес ефективнішим за допомогою штучного інтелекту, при цьому не турбуючись про перевитрати, безпеку даних і галюцинації.

Ця стаття для тих, хто вже «живе» в AWS і хоче працювати з ШІ всерйоз, а не на свій страх і ризик. Впізнаєте себе?

Фінансовий директор, який помітив, що відділ продажів платить за три різні ШІ-підписки з особистих карток, а в бухгалтерії ніхто не знає, що це за витрати.

Керівник служби безпеки, якому юрист зізнався, що «просто попросив чат-бота перевірити договір» — разом із реквізитами контрагента і сумою угоди.

Керівник клієнтського сервісу, чия команда щодня відповідає на сотні однакових звернень і не встигає до кінця зміни.

Технічний директор або тімлід, якого бізнес просить «вбудувати ШІ в продукт до кінця кварталу», але він не хоче створювати власну GPU-інфраструктуру і прив'язуватися до одного постачальника моделей.

Якщо так — далі ми розповімо, з чим зазвичай стикаються такі компанії і як це вирішує Claude від Anthropic в Amazon Bedrock.

З якими труднощами стикаються компанії

За нашими спостереженнями, ШІ сьогодні є майже в кожній компанії, але вбудувати його в реальні бізнес-процеси вдалося лише близько 10%. Заважають не самі моделі, а все, що навколо них.

«Тіньовий» ШІ. Люди не чекають на рішення IT-відділу і користуються тим, що знайшли самі. Компанія не знає, хто, що і куди відправляє.

Страх за дані. Головне питання, яке ми чуємо: «А наші договори не потраплять у навчання чужої моделі?» Поки на нього немає чіткої відповіді, юристи й безпека кажуть «ні».

Де фізично обробляються дані. Для українських і європейських компаній важливо, щоб інформація не виходила за межі ЄС і відповідала GDPR.

Немає контролю. Незрозуміло, хто має доступ, які запити надсилаються і як це потім перевірити під час аудиту.

Модель «вигадує». ШІ впевнено відповідає навіть тоді, коли не знає відповіді. Для внутрішньої довідки, юридичних чи фінансових питань це неприйнятно.

Оплата і договори. Закордонні сервіси вимагають окремих контрактів, валютних платежів і корпоративних карток. Бухгалтерія не в захваті.

Складно інтегрувати. Розробникам потрібен не «чат у браузері», а API, що витримує навантаження, легко підключається до існуючих систем і не прив'язує до однієї моделі.

Як це вирішує Amazon Bedrock

Amazon Bedrock — керований сервіс AWS, через який моделі Claude доступні всередині вашого облікового запису, без власних серверів і GPU. Ви платите лише за фактичне використання (**табл.**).

Таблиця. Проблеми, пов'язані з ШІ, та їх вирішення з Amazon Bedrock

Проблема	Як вирішує Bedrock
«Тіньовий» ШІ	Один корпоративний інструмент із доступом через IAM: компанія сама вирішує, хто і до яких моделей має доступ
Страх за дані	Запити і відповіді не передаються розробникам моделей і не використовуються для навчання; шифрування, ключі в AWS KMS
Де обробляються дані	Європейський профіль інференсу: запити обробляються лише в регіонах AWS у ЄС (Франкфурт, Ірландія, Париж, Стокгольм тощо)
Немає контролю	Кожен запит у AWS CloudTrail; приватне підключення без публічного інтернету через AWS PrivateLink
Модель «вигадує»	Граундинг через Knowledge Bases і перевірки Guardrails (детальніше нижче)
Оплата і договори	Витрати в звичному рахунку AWS, видно в Cost Explorer; через Elcore UA — розрахунки в локальній валюті
Складно інтегрувати	Єдиний API для всіх моделей: можна змінити модель без перебудови системи

Відповіді, які спираються на ваші документи

Граундинг (grounding) — це «прив'язка» відповідей моделі до перевірених джерел компанії. Amazon Bedrock Knowledge Bases індексують документи з Amazon S3, SharePoint, Confluence, Salesforce та інших сховищ. Claude відповідає на основі знайдених фрагментів і показує, звідки взято інформацію.

Додатково Guardrails перевіряють, чи підтверджена відповідь документами і чи стосується вона запитання, а слабкі відповіді відсікають. Вони ж блокують неприйнятний контент і приховують персональні дані. Для завдань, де

потрібна формальна точність — наприклад, перевірки відповідності політикам, — є Automated Reasoning checks.

Модель, налаштована під ваш бізнес

Тонке налаштування (fine-tuning) — це донавчання моделі на прикладах компанії: термінології, форматі відповідей, правилах класифікації звернень. У Bedrock навчальні дані залишаються у вашому Amazon S3, а налаштована модель приватна і доступна лише вам.

Для Claude донавчання доступне для окремих моделей і регіонів. Тому зазвичай ми починаємо з граундингу і роботи з промптами, а fine-tuning розглядаємо, коли потрібна стабільна поведінка на великих обсягах однотипних задач. Ще один варіант — Model Distillation: перенесення якості великої моделі на меншу, швидшу й дешевшу.

Модель під кожне завдання

- **Opus** — найпотужніша модель для складної аналітики, довгих документів і розроблення ПЗ. Найновіша версія, Claude Opus 5.5, з'явилася в AWS у вересні 2026 року.
- **Sonnet** — баланс якості, швидкості та ціни для більшості бізнес-сценаріїв.
- **Haiku** — швидка і економна модель для великих обсягів простих завдань.

Типова практика — це комбінування, де складне «віддають» Opus, а масове — Haiku.

Як це виглядає в житті

Найкраще Claude працює там, де люди витрачають години на читання, пошук і підготовку текстів.

Новий співробітник більше не смикає колег. Корпоративний асистент на базі Knowledge Bases відповідає на питання «як оформити відрядження?» чи «які умови в договорі з постачальником X?» і дає посилання на документ.

Юрист отримує чернетку за лічені хвилини, а не за день. Claude аналізує договір чи тендерну документацію, знаходить ризикові умови, порівнює редакції. Рішення приймає людина.

Оператор не тоне в однакових зверненнях. Claude сортує листи й заявки, пропонує чернетки відповідей у фірмовому стилі. Для великих обсягів добре підходить економна Haiku.

Агент закриває рутину сам. Створює заявку в CRM, звіряє рахунки, збирає звіт із кількох систем. Для таких сценаріїв AWS пропонує Amazon Bedrock AgentCore з контролем доступу і моніторингом.

Що будують розробники через API

Команди розробки вбудовують Claude у власні продукти через API Bedrock і стандартні SDK (Python, JavaScript, Java, .NET та інші). Найчастіше це:

- витягування даних — рахунки, накладні, анкети, скани й PDF перетворюються на JSON для ERP чи CRM;
- виклик інструментів (tool use) — модель сама звертається до ваших API і баз даних;
- пакетна обробка — batch inference для архівів листів, каталогів і тисяч звернень дешевше за звичайні запити;
- семантичний пошук — разом із моделями ембедингів у Bedrock (Amazon Titan Embeddings, Cohere Embed) пошук іде за змістом, а не за ключовими словами;
- переклад і локалізація — інтерфейси, документація й картки товарів кількома мовами;
- автоматизація в CI/CD — рев'ю pull request, генерація тестів, аналіз логів та інцидентів;
- розроблення з Claude Code — інструмент налаштовується на роботу через Bedrock, тож код не виходить за межі вашого AWS.

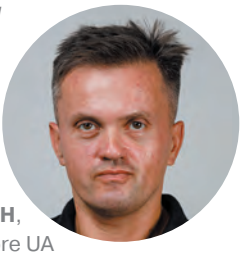
Щоб тримати витрати під контролем, Bedrock підтримує кешування промптів для повторюваного контексту, а прості запити легко спрямовувати на дешевші моделі.

З чого почати

1. Оберіть один процес, де ефект легко виміряти: скільки годин зараз іде на звернення, договори чи звіти.
2. Домовтеся про правила: які дані можна передавати моделі, хто має доступ, у якому регіоні обробляти.
3. Запустіть пілот на реальних даних з обмеженою групою користувачів і чіткими критеріями успіху.
4. Масштабуйте те, що спрацювало, і підберіть оптимальне поєднання моделей за ціною та якістю.

Як допомагає Elcore UA

> *Більшість компаній уже користується ШІ, проте мало хто знає, де саме обробляються їхні дані. Claude через Amazon Bedrock дає ту саму якість роботи, але в межах вашого облікового запису AWS.*



Олександр ФОМІН,
AWS Team Lead, Elcore UA

Як партнер AWS, Elcore UA супроводжує компанії від вибору першого сценарію і налаштування безпеки до запуску й подальшої підтримки. Тестування рішення можливе за рахунок ресурсів і маркетингових фондів AWS.

Що отримують клієнти:
- супровід фінансування пілотних проєктів;
- 10 годин безкоштовних консультацій щомісяця;
- розрахунки в локальній валюті згідно з договором;
- 25 днів відстрочки платежу;
- підтримку наших AI– та ML-фахівців після запуску.

elcore UA

Щоб обговорити ваш сценарій і запланувати пілот, пишіть команді AWS Elcore UA: aws@elcoregroup.com