

Технології штучного інтелекту та безпека даних

Освітня програма «Економіка і аналітика», 1 курс · 4 кредити ECTS

Лекція 10. Кібербезпека, посилена ШІ: фішинг, дипфейки, соціальна інженерія

Мета лекції: показати, як генеративний ШІ змінив масштаб, якість і доступність кіберзагроз — від фішингу до голосового клонування й дипфейків, — і дати конкретні технічні й організаційні методи захисту.

Зв'язок з попередніми лекціями: лекція 9 завершила блок про право тезою «ШІ — інструмент, відповідальність за людиною». Ця лекція показує іншу сторону тієї самої монети: ті самі інструменти, що автоматизують бізнес (лекція 6), так само активно використовуються зловмисниками проти цього бізнесу.

1. Місток від права до кібербезпеки

- Лекція 9 завершила блок про право й відповідальність тезою: ШІ — інструмент, відповідальність залишається за людиною.
- Ця лекція показує іншу сторону тієї самої монети: ті самі інструменти, які допомагають бізнесу автоматизувати рутину (лекція 6), так само активно використовуються зловмисниками для атак на цей самий бізнес.
- Питання лекції: як саме ШІ змінив ландшафт кіберзагроз, і які практичні заходи захисту існують — технічні та організаційні.

2. Чому ШІ змінює ландшафт кіберзагроз

- Пор. лекцію 8 (дипфейки): та сама технологія, що генерує корисний контент, генерує і переконливі фейки; та сама здатність писати «гарний людський текст» (лекція 2) робить фішингові листи невідрізними від легітимних.

Три ключові зміни, які ШІ вносить у кіберзагрози:

- **Масштаб** — раніше атака на 10 000 жертв вимагала 10 000 персоналізованих зусиль; тепер ШІ може згенерувати 10 000 персоналізованих повідомлень за секунди.
- **Якість** — зникають класичні маркери шахрайства (граматичні помилки, незграбні формулювання); текст, згенерований ШІ, часто якісніший за написаний нашвидкуруч людиною.
- **Доступність** — інструменти соціальної інженерії, які раніше вимагали спеціалізованих навичок, тепер доступні будь-кому за кілька кліків.

3. Фішинг до і після ШІ

До ШІ	Після ШІ
Грамматичні помилки, дивне форматування	Бездоганно грамотний текст
Неприродні звертання («Шановний клієнте»)	Персоналізація під конкретного отримувача

До ШІ	Після ШІ
Підозрілі посилання з очевидними помилками	Точна імітація корпоративного тону компанії
Легко впізнати за навченими правилами	Класичні маркери шахрайства зникають

Практичний наслідок для навчання безпеці: правило «шукай помилки в тексті» застаріло і більше не є надійним індикатором фішингу.

4. Соціальна інженерія, посилена ШІ

- Соціальна інженерія — маніпуляція людьми для отримання доступу до інформації чи систем, обходячи технічний захист.
- ШІ дозволяє автоматизувати збір і аналіз публічно доступної інформації про ціль (соціальні мережі, корпоративні сайти, новини) для створення високоперсоналізованого «приводу» (pretext) для атаки.

Приклад: замість загального «Шановний клієнте», зломисник може згенерувати лист, що згадує реальне ім'я керівника жертви, недавню новину про компанію і навіть стиль спілкування, типовий для галузі жертви — все це автоматично зібране й скомпоноване ШІ.

5. Голосове клонування та дипфейк-відео

- **Голосове клонування:** сучасні технології синтезу мовлення здатні відтворити голос конкретної людини на основі короткого аудіозразка — за деякими оцінками, кількох секунд запису (наприклад, з публічного відео) може бути достатньо для базового клонування.
- Практичний ризик: дзвінок, що звучить голосом реального керівника компанії з терміновим проханням здійснити переказ коштів чи розкрити конфіденційну інформацію.
- **Дипфейк-відео:** накладає обличчя чи міміку однієї людини на відео іншої, чи повністю генерує реалістичне відео неіснуючої події.
- Якість технології стрімко зростає (пор. лекцію 2 — вибухова еволюція якості генеративних моделей); ознаки, які раніше видавали дипфейк (неприродне моргання, дивні тіні), поступово зникають.

***Дидактична примітка:** Це прямий розвиток теми дипфейків з лекції 8 (та сама технологія генерує і корисний, і шкідливий контент), застосований конкретно до фінансового шахрайства.*

6. ВЕС-шахрайство: класичний сценарій, посилений ШІ

- **ВЕС (Business Email Compromise)** — шахрайство, за якого зломисник видає себе за керівника чи довірену особу компанії, щоб змусити співробітника здійснити фінансову операцію чи розкрити дані.
- **Класичний сценарій:** лист нібито від CEO з терміновим і «конфіденційним» проханням про переказ коштів, що обходить звичайні процедури погодження через тиск терміновості й авторитету.

- **Посилений ШІ сценарій додає:** бездоганно написаний лист у стилі конкретного керівника (на основі публічних виступів чи попереднього листування), а в найсерйозніших випадках — підтверджувальний телефонний дзвінок чи навіть відеодзвінок з клонованим голосом чи дипфейк-відео «керівника».

Це не гіпотетичний сценарій, а задокументована й зростаюча категорія фінансових злочинів у корпоративному середовищі по всьому світу.

7. Чому людина залишається найслабшою ланкою

- Технічні системи захисту (фаєрволи, антивіруси, спам-фільтри) стають дедалі досконалішими, тому зловмисники дедалі частіше атакують не технологію, а людську психологію.
- ШІ-посилена соціальна інженерія експлуатує ті самі психологічні принципи з лекції 4: **авторитет** (лист «від керівника»), **терміновість** (тиск на швидке рішення без перевірки), **соціальний доказ** (посилання на «інших колег, які вже погодили»).

Наскрізний зв'язок: пор. лекцію 4, п. 3 — «впевнений тон переконує» людей так само, як і те, що ми обговорювали про сприйняття відповідей ШІ; зловмисники свідомо використовують цю ж психологічну вразливість.

8. Технічні методи захисту

- **Автентифікація електронної пошти** (протоколи SPF, DKIM, DMARC) — технічні механізми, що ускладнюють підробку адреси відправника.
- **Багатофакторна автентифікація (MFA)** — навіть якщо пароль скомпрометовано, додатковий фактор захищає обліковий запис.
- **Технології виявлення аномалій** — системи, що аналізують типову поведінку користувача чи мережевого трафіку й сигналізують про відхилення.

Важливо: жоден технічний захід не є стовідсотковим — технічний захист має доповнюватися організаційними процедурами.

9. Організаційні методи захисту

- **Протокол зворотного дзвінка (call-back verification):** для будь-якого запиту на переказ коштів чи розкриття чутливої інформації — незалежно від того, наскільки терміновим і авторитетним він здається, — зателефонуйте відправнику за заздалегідь відомим (не тим, що вказаний у підозрілому повідомленні) номером для підтвердження.
- **Розподіл повноважень («принцип чотирьох очей»):** фінансові операції понад певну суму вимагають підтвердження щонайменше двох незалежних осіб.
- **Регулярне навчання співробітників** з актуальними прикладами ШІ-посилених атак, а не застарілими правилами «шукайте помилки в тексті».
- **Чіткі процедури ескалації** підозрілих запитів без страху «витратити чужий час» на перевірку.

10. Роль самого ШІ в захисті: та сама технологія працює і за захисників

- Дуальність технології (пор. лекцію 8) працює в обидва боки: ШІ використовується не тільки для атак, а й для захисту.
- Сучасні спам-фільтри й системи виявлення фішингу самі використовують машинне навчання для аналізу підозрілих патернів у листах.
- Системи виявлення шахрайства в банківській сфері (пор. лекцію 6) аналізують транзакції в реальному часі на предмет аномалій.

Практичний висновок: «гонка озброєнь» між атакувальним і захисним застосуванням ШІ — постійний процес, і жодна зі сторін не отримує остаточної переваги назавжди.

11. Чи можна надійно розпізнати ШІ-згенерований контент?

- **Технології водяних знаків (watermarking)** для згенерованого контенту — провайдери впроваджують приховані технічні маркери в зображеннях, аудіо, а іноді і в тексті, щоб алгоритмічно підтвердити ШІ-походження.
- **Обмеження:** водяні знаки можна видалити чи обійти технічно підкованому зловмиснику; стандарти ще не універсальні між провайдерами; для тексту технологія особливо ненадійна.

Практичний висновок: не покладайтесь виключно на технічні інструменти «виявлення ШІ» — вони недосконалі в обидва боки (пропускають згенероване, хибно позначають людське).

12. Практичний чек-лист розпізнавання підозрілого контенту

- Чи є в повідомленні елемент терміновості чи тиску ухвалити рішення негайно, без перевірки?
- Чи запит стосується грошей, доступу до системи чи розкриття чутливої інформації?
- Чи можу я перевірити особу відправника незалежним каналом (не тим, яким надійшов запит)?
- Чи узгоджується запит із типовою поведінкою цієї людини чи звичайними процедурами компанії?
- Якщо це голосове чи відеоповідомлення — чи можу я поставити несподіване контрольне питання, відповідь на яке важко підготувати заздалегідь?

13. Економічний вимір кіберзлочинності, посиленої ШІ

- Кіберзлочинність, посилена ШІ, — це не лише технічна, а й економічна проблема, що прямо стосується профілю цього курсу.
- **Пряма шкода:** втрачені кошти від успішних ВЕС-шахрайств можуть вимірюватися сотнями тисяч і мільйонами доларів для однієї компанії за один інцидент — на відміну від традиційного фішингу, що часто орієнтований на дрібніші суми масовим охопленням.

- **Непряма шкода:** витрати на розслідування інциденту, юридичний супровід, повідомлення постраждалих клієнтів (пор. лекцію 7 — юридичний обов'язок повідомити про витік), зростання страхових премій на кіберризики після інциденту.
- **Витрати на профілактику:** інвестиції в технічний захист, навчання персоналу та організаційні процедури (лекції 9–10 цієї теми) — це операційні витрати, які варто закладати в бюджет компанії так само системно, як витрати на фізичну безпеку офісу.

Практичний висновок для майбутнього економіста чи аналітика: оцінка кіберризiku дедалі частіше стає частиною звичайного фінансового планування й оцінки інвестиційних проєктів — а не вузькою темою виключно ІТ-відділу.

14. Наскрізнi зв'язки з попередніми лекціями

- **Лекція 2 (якість генерації)** — та сама здатність писати переконливий людський текст, що робить ШІ корисним, робить і фішинг небезпечнішим.
- **Лекція 4 (психологія переконливості, дипфейки)** — технічна основа для соціальної інженерії й дезінформації.
- **Лекція 8 (дуальність технології)** — та сама технологія генерує і корисний, і шкідливий контент.
- **Лекція 9 (відповідальність)** — організація, що стала жертвою BEC-шахрайства через недотримання процедур, також несе частину відповідальності за неналежний внутрішній контроль.

Місток до наступної лекції: ми розглянули, як генеративний ШІ змінив масштаб, якість і доступність кіберзагроз — від фішингу до голосового клонування й дипфейків — і які технічні й організаційні заходи захисту існують. Наступна лекція переходить від індивідуальних загроз до організаційного рівня: як компанія системно впроваджує безпечне використання ШІ — політики, ролі, відповідальність. Це тема лекції 11.

Контрольні питання для самоперевірки студентів

- Поясніть, чому правило «шукайте граматичні помилки» більше не є надійним індикатором фішингу.
- Наведіть приклад того, як ШІ дозволяє персоналізувати атаку соціальної інженерії у масштабі.
- Опишіть класичний сценарій BEC-шахрайства і поясніть, як ШІ його підсилює.
- Що таке протокол зворотного дзвінка (call-back verification) і чому він ефективний навіть проти дуже переконливих дипфейків?
- Поясніть своїми словами, чому «гонка озброєнь» між атакувальним і захисним застосуванням ШІ навряд чи колись завершиться остаточною перемогою однієї зі сторін.
- Складіть свій приклад контрольного питання, яке важко підготувати заздалегідь зловмиснику, що використовує клонований голос колеги.