

ОҚЫТУ КУРСЫ

# Жасанды интеллект заманындағы фейктер

*Қалай алданбауға болады?*

Deerfake, фейк жаңалықтар, манипуляция — бүгінгі  
цифрлық қауіптерден қорғанудың толық нұсқаулығы



# Не үйренесіз?

01

## Фейктің түрлері

Мәтін, аудио, видео, сурет фейктері

02

## AI технологиялары

Deerfake, GAN, LLM қалай жұмыс жасайды

03

## Тану әдістері

Визуалды белгілер мен тексеру жолдары

04

## Тексеру құралдары

Факт-чекинг сайттары мен сервистері

05

## Психология

Неліктен алданамыз? Когнитивті қателіктер

06

## Қорғану стратегиясы

Жеке және ұйымдық деңгейдегі қорғаныс

# Фейк дегеніміз не?

*Фейк — адамдарды жаңылыстыру немесе зиян келтіру мақсатында жасалған жалған ақпарат немесе өзгертілген медиа мазмұн.*



## Мәтіндік фейк

- Жалған жаңалықтар
- Дезинформация
- Пропаганда
- Астыртын жарнама



## Медиа фейк

- Deepfake видео
- AI-генерленген суреттер
- Клондалған дауыс
- Синтетикалық аудио



## Контекст фейк

- Ескі фото — жаңа оқиға
- Сатира — шынайы ретінде
- Жартылай шындық
- Манипуляциялық монтаж

# AI қалай фейк жасайды?

## Deepfake (GANs)

Генеративті-адверсариалды желілер екі нейронды желіні бір-біріне қарсы оқытады. Нәтижесінде шынайы адамның бет-әлпетін басқа дене үстіне «кигізеді».

## LLM (GPT т.б.)

Үлкен тілдік модельдер адам сияқты мәтін жазады. Жалған мақалалар, хаттар, пікірлер автоматты түрде миллиондап жасала алады.

## Voice Cloning

3–10 секунд аудио үлгісінен бастап кез келген адамның дауысын клондап, бейтаныс мазмұн айтқызуға болады.

## Статистика

500%

deepfake өсімі  
2023–2024 жж.

96%

deepfake  
порнография

\$25M

бір компания  
шығыны (2024)

# Deepfake-ті қалай тануға болады?



## Жыпылықтау

Табиғи жыпылықтаудың жоқтығы немесе артықшылығы



## Жарық

Беттегі жарық мен фондағы жарық сәйкес келмейді



## Шекаралар

Шаш, құлақ, мойын айналасындағы бұлыңғырлық



## Эмоция

Бет мимикасы мен дауыс интонациясы сәйкессіз



## Ерін синхроны

Аудио мен ерін қозғалысы бір-біріне сай емес



## Қолдар

AI-ға қол саусақтарын дұрыс жасау қиын



## Тіс/көз

Тістер бұлыңғыр, көз жарығы дұрыс емес




## Текстура

Тері текстурасы тым тегіс немесе пластик сияқты

# Жалған жаңалықтың белгілері

 Тақырыпта CAPS LOCK, «шок!», «үшін!» деген сөздер

 Беймәлім домен: .xyz, .biz, сайт атауы ресми сайтқа ұқсас

 Дата жоқ немесе ескі оқиғаны жаңа ретінде ұсынады

 Авторы жоқ, тек «редакция» немесе бейтаныс есім

 Сілтемелер тексерілмейді немесе мүлдем жоқ

 Мазмұн шамадан тыс эмоционалды, ашу-ыза тудырады

## SIFT әдісі

**S**

### Stop

Тоқтаңыз. Реакция бермес бұрын ойланыңыз

**I**

### Investigate

Дереккөзді тексеріңіз

**F**

### Find Coverage

Басқа сенімді сайттардан іздеңіз

**T**

### Trace Claims

Мәліметтің түп-тамырына жетіңіз

# Тексеру құралдары

## Факт-чекинг сайттары

**FactCheck.org**

АҚШ саясаты мен ғаламдық мәселелер

**Snopes.com**

Аңыздар мен вирусты жалған ақпараттар

**PolitiFact**

Саясаткерлер сөздерін тексеру

## Сурет тексеру

**Google Images**

Кері сурет іздеу (reverse search)

**TinEye**

Суреттің шыққан жері мен тарихы

**FotoForensics**

Суреттің өзгертілгенін анықтау

## Deepfake детекторлар

**Microsoft Video Authenticator**

Видео аутентификациясы

**Deerware Scanner**

Тегін deepfake сканері

**Sensity AI**

Кәсіби деңгейдегі детектор

# Суретті қалай тексеруге болады?

1

## Суретке оң жаққа басыңыз

Chrome браузерінде суретке оң жаққа басып, «Google-дан суретті іздеу» таңдаңыз.

2

## TinEye-ға жүктеңіз

tineye.com сайтына кіріп, суретті жүктеңіз немесе URL кірістіріңіз.

3

## Нәтижелерді талдаңыз

Сурет бұрын қашан, қайда жарияланғанын тексеріңіз. Ескі сурет жаңа оқиғада қолданылып жатқан болуы мүмкін.

4

## Метадеректерді тексеріңіз

Jeffrey's Exif Viewer арқылы суреттің геолокациясы, камерасы, уақыты туралы ақпаратты алыңыз.



# Неліктен алданамыз?

## Растау жаңылысы

Өз пікірімізді растайтын ақпаратқа сенуге бейімбіз

## Жалғандық эффекті

Бір ақпаратты жиі көрген сайын, оны шын деп қабылдаймыз

## Эмоционалды ойлау

Қатты эмоция тудырған ақпаратты логикасыз бірден бөлісеміз

## Бедел жаңылысы

Танымал адам айтса, тексермей сенеміз

## Сыни ойлаудың 3 сұрағы

1

**КІМ жариялады?**

Авторды, сайтты тексеріңіз

2

**НЕГЕ маған бұл жіберілді?**

Мақсатты ойланыңыз

3

**ҚАЙДА тексеруге болады?**

2-3 тәуелсіз дереккөзді іздеңіз

# Әлеуметтік желіде сақ болыңыз

⚠ Вирусты мазмұнның 70%-дан астамы адамдар тексермей-ақ бөліседі!



## WhatsApp / Telegram

- «Пересылка» деген хабарды сақтықпен оқыңыз
- Паника тудыратын хабарламаларды тексермей бөліспеңіз
- Жіберуші санағышы — сенімсіздік белгісі



## Instagram / TikTok

- Тренд бейнелерді критикалық көзбен қараңыз
- Коментарийлердегі дезинформацияға назар аударыңыз
- Аккаунттың жасы мен жазбалар тарихын тексеріңіз



## Facebook / X (Twitter)

- Репост алдында дереккөзге өтіңіз
- Боттардың белгілерін тануды үйреніңіз
- Сатиралық аккаунт белгілерін ескеріңіз

# Дауыс клонингі — жаңа қауіп

НАҚТЫ ЖАҒДАЙ

2024 жылы Гонконгда компания қызметкері директордың клондалған дауысы арқылы өткізілген видео конференцияда \$25 миллион аударды. Барлық қатысушылар deerfake болды.

## Дауыс клонингінің белгілері

- › Дауыстағы жасанды, металл реңк
- › Фондық шу мүлдем жоқ
- › Сөз тоқтамалары табиғи емес
- › Аты-жөніңізді дауысты арнайы айтады
- › Шұғыл төлем немесе деректер сұрайды

## Қорғану жолдары

- › Отбасыңызбен «кодтық сөз» белгілеңіз
- › Қайта хабарласыңыз — таныс нөмірге
- › Ешқашан бір қоңырауда ақша аудармаңыз
- › Банк хабарласса — өзіңіз банкке қоңырау шалыңыз
- › Интуицияға сеніңіз — бір нәрсе дұрыс емес деп сезсеңіз

# Жеке қорғану стратегиясы

## 1-деңгей: АЛДЫН АЛУ

- ✓ Медиасауаттылықты үнемі арттырыңыз
- ✓ Цифрлық гигиенаны сақтаңыз
- ✓ Танымал фейк-чек сайттарын бетбелгіге қосыңыз

## 2-деңгей: ТЕКСЕРУ

- ✓ Бөліспес бұрын — тексеріңіз
- ✓ Эмоционалды реакция = сақтық белгісі
- ✓ SIFT әдісін қолданыңыз

## 3-деңгей: ХАБАРЛАУ

- ✓ Фейкті платформаға хабарлаңыз
- ✓ Жакындарыңызды ескертіңіз
- ✓ Фейк туралы ашық айтыңыз



Есте сақтаңыз: Фейкті бөлісу — оны жасаушымен бірге жауапты болу.

# Ұйымдық деңгейдегі қорғаныс

## Корпоративтік саясат

- › Медиамаaterialдарды тексеру нұсқаулығы жасаңыз
- › Дерекпен жұмыс жасау стандарттарын белгілеңіз
- › Инцидент туралы есеп беру процесін құрыңыз

## Командалық оқыту

- › Тұрақты медиасауаттылық жаттығулары
- › Фишинг симуляциялары
- › Deepfake хабардар ету сессиялары

## Техникалық шаралар

- › Deepfake детектор бағдарламалары
- › Дауысты растау жүйелері
- › Мультифакторлы аутентификация

## Дағдарыс жоспары

- › Жалған ақпарат тарады дегенде не істейсіз?
- › Жауапты адам белгіленген болуы тиіс
- › ЖАҚ-қа хабарлау алгоритмі дайын болсын

# Өзіңізді тексеріңіз!

1. Сіз WhatsApp-та «ШҰҒЫЛ! Вакцина зиянды!» деген хабарлама алдыңыз. Не істейсіз?

A. Дереу барлық топтарға жіберемін

B. Авторын тексеріп, фактчек сайтынан іздеймін ✓

C. Маған жіберілсе, шынайы шығар

D. Жоқ, назар аудармаймын

2. Видеода таныс саясаткер жанжалды нәрсе айтып тұр. Бірінші қадамыңыз?

A. Бөлісемін, өйткені мен оны жақтырмаймын

B. Deepfake детектормен тексеремін

C. Ресми дереккөздерден растаймын ✓

D. B және C дұрыс

3. SIFT моделіндегі «S» нені білдіреді?

A. Search — іздеу

B. Stop — тоқтау ✓

C. Share — бөлісу

D. Source — дереккөз

Қ О Р Ы Т Ы Н Д Ы

# Сіз дайынсыз!

01

Кез келген ақпаратқа сенімсізлікпен қараңыз — бұл дұрыс позиция

02

SIFT әдісін қолданыңыз: тоқтаңыз, тексеріңіз, растаңыз

03

Deerfake белгілерін біліңіз: жарық, ерін синхоны, шекаралар

04

Тексеру құралдарын бетбелгіге сақтаңыз және үнемі қолданыңыз

05

Эмоция жоғарыласа — тексеру сигналы деп есептеңіз



Білім — ең жақсы қорғаныс. Бүгіннен бастап медиасауаттылықты өмір салтына айналдырыңыз!