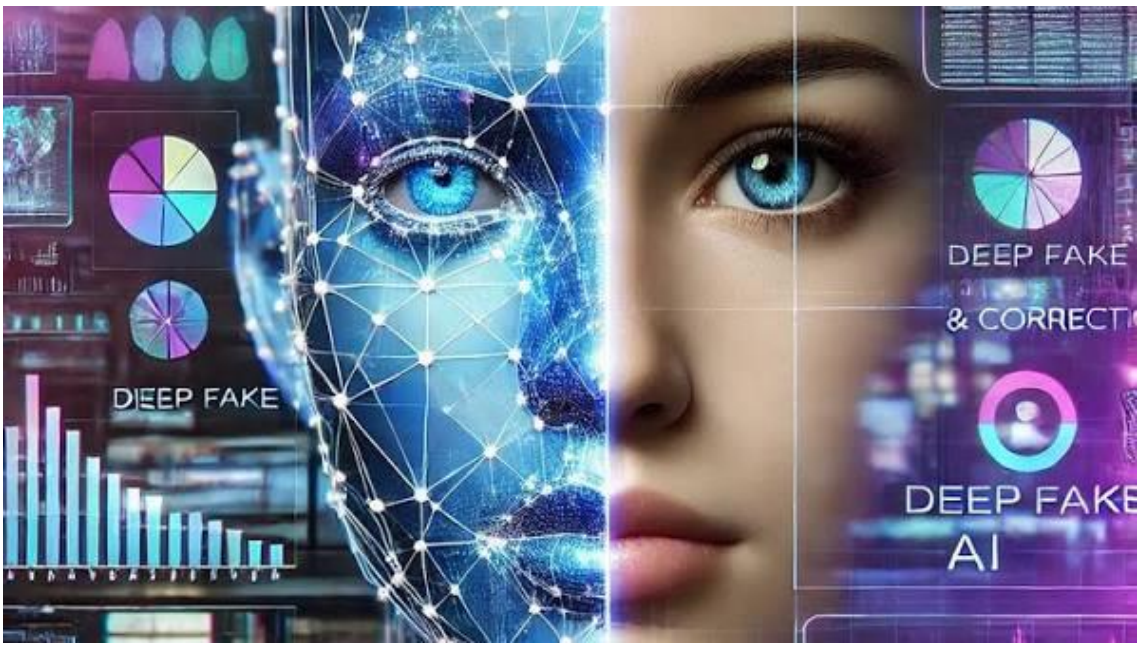
Дипфейк: шындыққа сену қиындаған дәуір

Әлеуметтік желіден белгілі бір адамның мәлімдеме жасап тұрған видеосын көрдіңіз делік. Оның бет-әлпеті анық көрінеді, дауысы өзінікі сияқты естіледі, ерін қимылы айтылған сөздермен сәйкес келеді. Видеоның сапасынан да күмән тудыратын ештеңе байқалмайды. Бір қарағанда, оның жалған екеніне сену қиын. Бірақ видеодағы адам бұл сөздерді өмірінде ешқашан айтпаған болуы мүмкін.

Бірнеше жыл бұрын мұндай жағдай фантастикалық фильмнің сюжеті сияқты көрінетін. Бүгінде жасанды интеллекттің дамуы адамның бейнесін, дауысын және іс-әрекетін цифрлық түрде қайталап, шынайыға өте ұқсас контент жасауға мүмкіндік беріп отыр. Осындай технологиялардың нәтижесінде **«дипфейк»** ұғымы пайда болды.



Жасанды интеллект дамыған сайын шынайы және жасанды контентті ажырату қиындап келеді. umd.edu

• ***Дипфейк дегеніміз не?***

«Deepfake» сөзі ағылшын тіліндегі "deep learning" - «терең оқыту» және "fake" - «жалған» сөздерінің бірігуінен шыққан. Қарапайым тілмен айтқанда, дипфейк - жасанды интеллект көмегімен адамның бейнесін, дауысын немесе әрекетін өзгертіп, шынайы болып көрінетін жасанды контент жасау технологиясы.

Мысалы, адамның жүзін басқа адамның денесіне орналастыруға, оның дауысын қайталауға немесе ешқашан айтпаған сөздерін айтқызып көрсетуге болады. Бұрын мұндай жұмысты жасау үшін кәсіби мамандар, арнайы

бағдарламалар және көп уақыт қажет болса, қазір генеративті жасанды интеллект бұл процесті айтарлықтай жеңілдетті.

Дипфейктің өзі міндетті түрде зиянды технология емес. Оны кино өндірісінде, дубляжда, жарнамада, білім беруде және шығармашылық жобаларда пайдалануға болады. Мәселе технологияның қандай мақсатта қолданылатынында.

• Жасанды видеолар неге қауіпті болып барады?

Алғашқы дипфейктерді анықтау салыстырмалы түрде оңай болатын. Кейіпкердің беті табиғи қозғалмайтын, көзқарасы өзгеріп кететін, жарық пен көлеңке сәйкес келмейтін немесе ерін қозғалысы дауысқа дәл келмейтін жағдайлар кездесетін.

Алайда жасанды интеллект өте жылдам дамып келеді. Қазіргі генеративті модельдер тек адамның жүзін ауыстырумен шектелмейді. Олар мәтін бойынша тұтас көрініс, кейіпкерлер, қозғалыс, қоршаған орта және дыбыс жасай алады.

Дипфейк технологиясы адамның бет-бейнесі мен дауысын жасанды интеллект көмегімен өзгертуге мүмкіндік береді. technologyreview.com

2026 жылы шыққан **"Alibaba Cloud Wan 3.0"** моделі осы дамудың бір мысалы болып табылады. Ол мәтін, сурет, аудио және видео сияқты бірнеше материалды пайдаланып, бір генерацияда 30 секундқа дейінгі, 1080p сапасындағы видео жасай алады. Модель диалог, фондық музыка және дыбыс эффектілерін де видеоның өзімен бірге генерациялай алады. Сонымен қатар

кейіпкердің бет әлпеті мен басқа ерекшеліктерін кадрлар арасында тұрақты сақтауға бағытталған мүмкіндіктері бар.

Wan 3.0 дипфейк жасау үшін арнайы шығарылған құрал емес. Alibaba Cloud оны кино, жарнама, ойын индустриясы және контент өндірісі сияқты шығармашылық бағыттарға арналған видео генерациялау технологиясы ретінде ұсынады. Бірақ осындай модельдердің мүмкіндіктері бір маңызды өзгерісті көрсетеді: компьютер арқылы жасалған видео барған сайын шынайы болып келеді.

Бүгінде мәселе тек адамның бетін басқа адамның бетімен ауыстыруда емес. Жасанды интеллект өмірде мүлдем болмаған оқиғаның өзін видео түрінде жасап бере алады.

• «Өз көзіммен көрдім» деген дәлел жеткілікті ме?

Адамдар ұзақ уақыт бойы фото мен видеоға оқиғаның дәлелі ретінде қарады. Біреу қандай да бір жағдай туралы айтса, *«видеосы бар ма?»* деген сұрақ қою қалыпты нәрсе.

Бірақ генеративті жасанды интеллект дамыған сайын бұл түсінік өзгеріп келеді.

Егер жасанды интеллект болмаған оқиғаның видеосын жасай алса, онда видеоның болуы оның міндетті түрде болғанын білдірмейді. Сондықтан болашақта *«мен өз көзіммен видеодан көрдім»* деген сөз ақпараттың шынайылығын дәлелдеуге жеткіліксіз болуы мүмкін.

Бұл әсіресе әлеуметтік желілерде маңызды. Адамдар көбіне видеоны толық тексерместен бөліседі. Эмоция тудыратын видео бірнеше сағаттың ішінде мыңдаған немесе миллиондаған адамға тарауы мүмкін. Ал оның жалған екені кейін анықталса да, алғашқы ақпараттың әсерін толығымен жою қиын.

UNESCO да қазіргі ақпараттық кеңістікте жасанды интеллект арқылы жасалған контент пен дипфейктер адамдардың шындықты жалғаннан ажырату қабілетіне жаңа сынақ туғызатынын атап өтеді. Ұйымның пікірінше, ақпараттың көп болуы оның сапалы немесе шынайы екенін білдірмейді.

• Дипфейк және журналистика

Дипфейктің дамуы журналистикаға ерекше әсер етеді. Себебі журналистиканың негізгі құндылықтарының бірі - шынайылық.

Журналист әлеуметтік желіден маңызды оқиғаның видеосын көрді делік. Бұрын оның қай жерде және қашан түсірілгенін анықтау негізгі міндеттердің бірі болса, енді тағы бір сұрақ пайда болды: *"бұл видеоның өзі шынайы ма?"*

Сондықтан журналист тек видеоны тауып, жариялаумен шектелмеуі керек. Материалдың бастапқы дереккөзін анықтау, оны алғаш кім жариялағанын тексеру, видеодағы адам немесе ұйыммен байланысу, басқа тәуелсіз дереккөздермен салыстыру және қажет болған жағдайда цифрлық тексеру құралдарын пайдалану маңызды.

Мысалы, әлеуметтік желіде танымал адамның даулы мәлімдеме жасап тұрған видеосы тараса, журналист оны бірден жаңалық ретінде жарияламауы керек. Алдымен видеоның бастапқы нұсқасын тауып, мәлімдеменің басқа жазбалары бар-жоғын және адамның ресми өкілдері не айтқанын тексерген жөн.

Жылдамдық журналистика үшін маңызды болғанымен, жасанды интеллект дәуірінде **дәлдік жылдамдықтан да маңызды болуы мүмкін.**

• Дипфейктің тағы бір қаупі - шынайы ақпаратқа сенбеу

Дипфейк мәселесінің қызық әрі қауіпті жағы бар. Оның қаупі адамдардың жалған видеоға сеніп қалуымен ғана шектелмейді.

Керісінше жағдай да болуы мүмкін.

Егер қоғам жасанды видеолардың бар екенін жақсы білсе, адамдар өздеріне ұнамайтын шынайы видеоны да *«бұл дипфейк»* деп жоққа шығара алады.

Осылайша екі мәселе қатар пайда болады: жалған ақпарат шынайы болып көрінуі мүмкін, ал шынайы ақпарат жалған деп қабылдануы мүмкін.

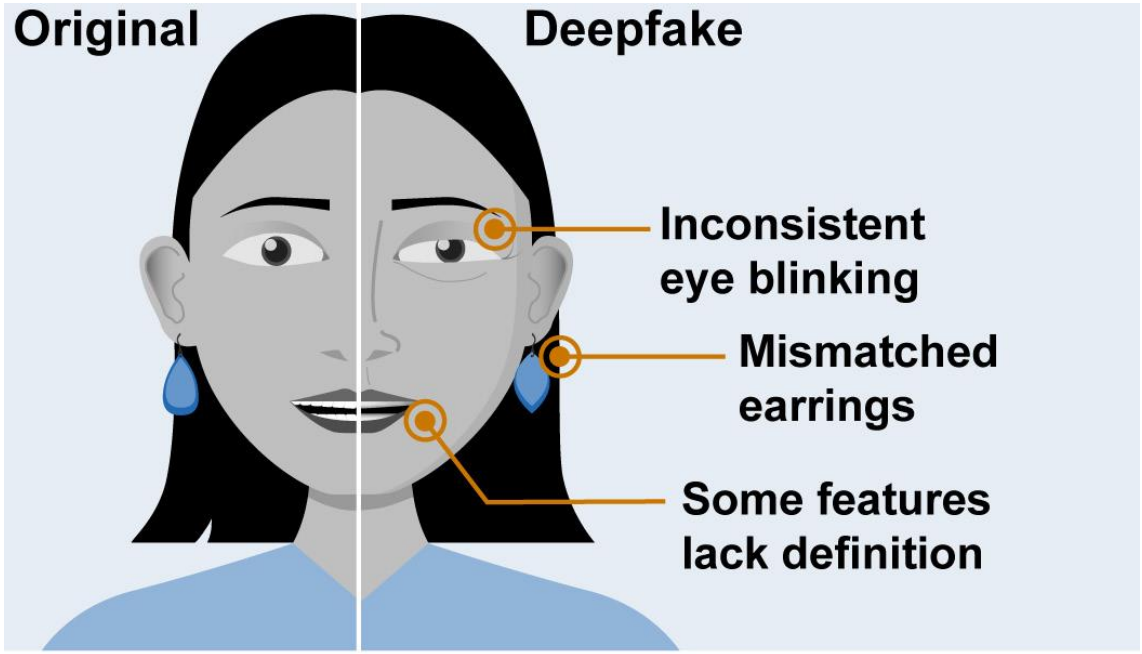
Нәтижесінде адамдардың видеоға, жаңалықтарға және жалпы ақпарат көздеріне деген сенімі төмендеуі ықтимал. Бұл журналистика үшін өте маңызды мәселе. Себебі БАҚ-тың жұмысы тек ақпарат тарату емес, аудиторияға сол ақпараттың қалай алынғанын және неге оған сенуге болатынын көрсетуге де байланысты.

• Дипфейкті қалай анықтауға болады?

Бұрын дипфейкті адамның көзінің дұрыс жыпылықтамауы, ерін қозғалысының сәйкес келмеуі немесе беттегі жарықтың өзгеруі сияқты белгілер арқылы байқауға болатын. Бірақ технология жетілген сайын тек визуалды белгілерге сену жеткіліксіз болып барады.

Сондықтан ең тиімді тәсілдердің бірі - **видеоны емес, оның шығу тарихын тексеру**.

Видео қай аккаунтта алғаш жарияланды? Оны сенімді ақпарат құралдары жариялады ма? Видеода көрсетілген оқиға туралы басқа дереккөздер не дейді? Толық нұсқасы бар ма? Видеодағы адам немесе ұйым бұл ақпаратты растады ма?



Source: GAO; conceived from DARPA image at https://www.darpa.mil/news-events/2019-09-03a. | GAO-20-379SP

Күмәнді видеоны тек сыртқы белгілері бойынша емес, оның бастапқы дереккөзі мен контексті арқылы да тексеру қажет. gao.gov

Осындай қарапайым сұрақтардың өзі жалған ақпараттың таралуын азайтуға көмектеседі.

Сонымен қатар журналистер үшін кері сурет іздеу, видеоның жеке кадрларын тексеру, файлдың метадеректерін зерттеу және фактчекинг құралдарын қолдану маңызды бола түседі.

Алайда ешбір автоматты «AI detector»-ды абсолютті дәл құрал деп қабылдауға болмайды. Жасанды интеллект дамыған сайын оны анықтайтын технологиялар да қатар дамиды. Сондықтан техникалық құрал журналистік тексерудің орнын баспайды, тек оған көмектеседі.

Біз ақпараттың жаңа кезеңіне кіріп келеміз. Бұрын мәтіндегі ақпаратқа күмәндансақ, кейін фотосуреттің өңделген-өңделмегенін тексеруді үйрендік. Енді алдымызда видео мен аудионың шынайылығын анықтау мәселесі тұр.

Wan 3.0 сияқты жаңа генеративті модельдер жасанды интеллекттің мүмкіндігі қаншалықты жылдам дамып жатқанын көрсетеді. Бүгін 30 секундтық шынайы көрінетін видео жасайтын технология ертең одан да ұзақ әрі күрделі контент жасай алады. Сондықтан технология дамыған сайын қоғамның медиасауаттылығы да қатар дамуы қажет.

Дипфейктің ең үлкен қаупі - жалған видеолардың пайда болуы ғана емес. Ең үлкен қауіп - **адамдардың ненің шын, ненің жалған екенін ажырата алмайтын жағдайға жетуі**.

Осындай жағдайда журналистиканың маңызы азаймайды. Керісінше, тексерілген ақпараттың құндылығы артады.

Жасанды интеллект шындықты жойып жатқан жоқ. Бірақ ол **шындықты дәлелдеуді қиындатып жатыр**. Сондықтан болашақ журналистің міндеті ақпаратты бірінші болып жариялау ғана емес, оның неліктен шынайы екендігін дәлелдей алу болуы мүмкін.

Бекмұханбет Дінмұхамед