



МЭДЭЭЛЛИЙН ШИНЭ ХЭРЭГСЭЛ ДЭХ СИНТЕТИК МЕДИА БУЮУ ГҮН ХУДАЛ КОНТЕНТИЙН СУДАЛГАА

Б.Чинзориг

**МУБИС-ийн сэтгүүл зүйн тэнхимийн багш, МСНЭ-ийн удирдах зөвлөлийн
 гишүүн, Доктор (Ph.D)**

ХИЙМЭЛ ОЮУН УХААН БА ХЭВЛЭЛ МЭДЭЭЛЭЛ

- Виртуал машин (Virtual Assistant Media anchor)
- WEB-4.0 буюу Симбиотик веб (Symbiotic web)



ХИЙМЭЛ ОЮУН УХААНТ МЭДЭЭЛЛИЙН ХЭРЭГСЛИЙГ ДОТОР НЬ

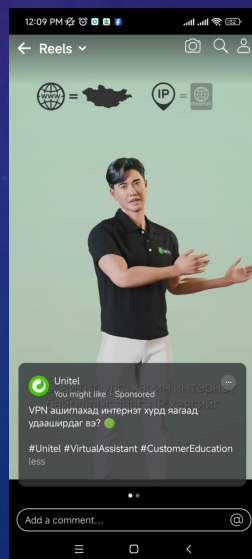
- Ерөнхий хиймэл оюун ухаант
- Бичил хиймэл оюун ухаант /Сэтгүүлчийн ажлыг хөнгөвчлөх программууд/

МАНАЙД



- Хязгаарлагдмал үйл ажиллагаа бүхий бичил хиймэл оюун ухаан /Voyager AI, Sara, Чимэгэ системс, Болор софт, Олон улсын платформууд/
- Олон улсад хөгжүүлж буй гүн суралцах чадвар бүхий ерөнхий хиймэл оюун ухааны ололт / Chat GPT IV/

ЧИМЭГЭ



ГҮН ХУДАЛ КОНТЕНТ БА МОНГОЛЫН МЭДЭЭЛЛИЙН ШИНЭ ХЭРЭГСЭЛ

- Хиймэл оюун ухааны орчин үеийн ололт амжилтууд зохиомол мэдээллийн цоо шинэ төрөл болох гүн худал контент (Deep Fake) гэх синтетик медиа (Synthetic media) хэлбэрээр бүтээсэн худал мэдээллийн цоо шинэ хэлбэрийг бий болгожээ.
- Аудио, видео манипуляц



Мөн гүн хуурамч контентийг бэлтгэдэг арга, техник хэрэгслүүд улам боловсронгуй хүртээмжтэй болж байна.

Yadlin-Segal, A. O. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research Into New Media Technologies*, 27(1), 36-51. doi:<https://doi.org/10.1177/1354856520923963>

Гүн худал контент бэлтгэж буй программууд нь хиймэл оюун ухааны машин сургалтын аргуудыг хэрэглэхийн зэрэгцээ бодит хүний видео симуляци эсвэл нүүр царайг солих байдлаар хуурамч дүрсийг боловсруулдаг, зураг бичлэгийг зохиомлоор бүтээдэг, мөн гүн хуурамч контентийг илрүүлээд алдааг олж, түүн дээрээ суралцдаг Generative Adversarial Networks (GANs) технологид тулгуурлан хөгжиж бие даан худал контентийг бэлтгэж чаддаг болжээ.

Тоглоомын онолын (Theory of game) математик тооцооллууд болох ICCT (In computational complexity theory) болон NP (non deterministic polynomial time) загваруудыг ашиглан дедуктив үндэслэл, индуктив үндэслэл, эрсдэлд суурилсан үнэлгээг ч гүйцэтгэх хэмжээнд хөгжжээ.

ГҮН ХУДАЛ КОНТЕНТИЙГ АНГИЛЖ ҮЗВЭЛ

- Агуулгын хувьд гүн худал контент нь жендэр, гарал үүсэл, үндэстэн ястнаар ялгаварлан гадуурхах, порнограф, нийгмийн халамж гэсэн сэдвүүдийг хамарч байгаа
- Бие мах бодийг өөрчилсөн, үг яриаг өөрчилсөн гэсэн хэлбэртэйгээр бэлтгэгдэж байна.

БИЕ МАХ БОДИЙГ ӨӨРЧИЛСӨН ГҮН ХУДАЛ КОНТЕНТ

- Олон улсын судалгаанд гүн хуурамч контентийн энэ хэлбэрийн дийлэнх буюу 96 хувийг зохиомол порнографууд эзэлж байсан талаар онцолжээ.

Van Der Nagel, E. (2020). Verifying images: Deepfakes, control, and consent. Porn Studies, (4), 424-429.
doi:<https://doi.org/10.1080/23268743.2020.1741434>.)

- Data miner программыг ашиглан дэлхий дахинд байгаа Interface веб хуудсуудад байгаа энэ төрлийн зураг дүрс бичлэгийн тоог гаргаж үзвэл сүүлийн 3 сарын хугацаанд 2.765.852 ширхэгийг бэлтгэж цахим орчинд түгээсэн байна. Эдгээрийн дийлэнх хувийг К поп хамтлагийн эмэгтэй дуучид болон Солонгос олон ангит киноны эмэгтэй жүжигчдийн бие махбодийг өөрчилсөн зураг, бичлэг эзэлж байлаа. Тэгвэл Монгол улсын цахим орчинд байгаа энэ төрлийн зураг бичлэгийг сүүлийн 5 жилээр судалгааг хийж үзвэл олон улсын түвшингөөс харьцангуй цөөн байна.



Манай улсад 2026 оны эхний улирлын байдлаар 229 ширхэг байгаа нь ажиглагдсан бөгөөд улс төрч эмэгтэйчүүд болон дуучин эмэгтэйчүүдийн видео, зургууд дийлэнх хувийг эзэлж байна. Ц. Оюунгэрэл, Б.Баярсайхан, Б.Саранчимэг, Г.Мөнхцэцэг, Нара, гэх мэт эмэгтэй улс төрчид болон А.Хишигдалай, Mrs M, Nene зэрэг эмэгтэй дуучидтай холбоотой гүн худал контентийн тоо бусад хүмүүсийнхээс харьцангуй олон ажиглагдав. Гэхдээ эдгээр нь харьцангуй цөөн хэрэглэгчдийн дунд тархсан байна.

Тэгвэл манайд энэ төрлийн гүн худал контентоос Эрчим хүчний сайд Д.Чойжилсүрэнгийн нүүрийг эмэгтэй хүний биед залгасан видео бичлэг хамгийн олон хандалт авчээ. Тус бичлэгийг нэр бүхий 13 мэдээллийн сайт, таван мянгаас дээш дагагч бүхий 94 пэйж хуудас, 100 орчим групп түгээсэн бөгөөд давтагдсан тоогоор 13762 сэтгэгдэл авч, 2 сая орчим удаа үзжээ. Уг бичлэг нь бодит гэсэн сэтгэгдэл төрүүлэх зорилгогүйгээр зүгээр шоглох агуулга бүхий гүн худал контент байв. Гэхдээ тус мэдээллийг хэвлэл мэдээллийн хэрэгслүүд болон сэтгүүлчид аль аль нь түгээж мэдээлсэн нь мэргэжлийн үүднээс авч үзвэл алдаа юм.

Судлаачийн дүгнэснээр Эр эм ямар ч хүний нүүр царайг сольж элдэв хэлбэр дүрстэй болгож шоглох нь хувь хүний ёс зүйн хувьд нийцэхгүй бөгөөд цаашилбал цахим гэмт хэргийн шинжтэй асуудал юм.

Gosse, C. B. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37 (5), 497-511. doi: <https://doi.org/10.1080/15295036.2020.1832697>)

ҮГ ЯРИАГ ӨӨРЧИЛСӨН ГҮН ХУДАЛ КОНТЕНТ:

- Энэ төрлийн контентийг олон улсад болон манай улсад тогтмол ашиглаж байна. Тухайлбал, 2022 оны 03-р сард хэд хэдэн хэвлэл мэдээллийн хэрэгслээр Украины ерөнхийлөгч Владимир Зеленский Орост бууж өгч буйг харуулсан гүн хуурамч бичлэгийг түгээснийг Мета болон YouTube олж устгажээ.

Wakefield, J. (2022 оны 3 18). Deepfake presidents used in Russia-Ukraine war. BBC.
<https://www.bbc.co.uk/news/technology-60780142>

- Бид Rapid Miner программыг ашиглан судалгааг хийж үзвэл тус контентийг нийтлэгдсэнээс хойш 180 минутын дотор эх сурвалжаас нь устгасан ч 27 хэлээр орчуулагдан дэлхий дахинд тархжээ. Манай монголын зарим мэдээллийн хэрэгсэл ч уг контентийг субтитр хийж түгээсэн байна.

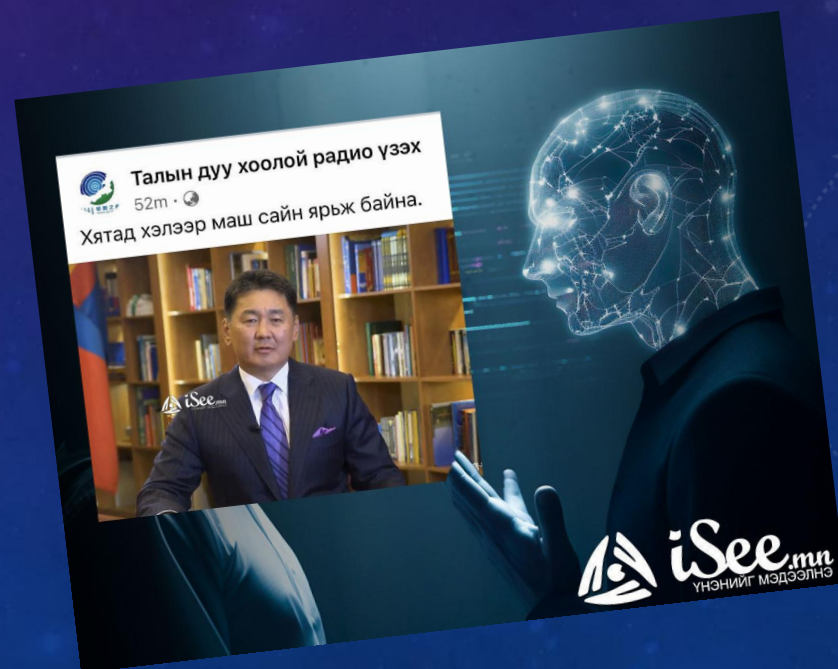


МАНАЙ УЛСЫН ХУВЬД ТӨРИЙН ӨНДӨР АЛБАН ТУШААЛТНУУДЫН ХИЙСЭН МЭДЭГДЭЛ ХЭЛСЭН ҮСГИЙГ ӨӨРЧИЛСӨН ГҮН ХУДАЛ КОНТЕНТ СҮҮЛИЙН ҮЕД ЦӨӨНГҮЙ АЖИГЛАГДДАГ БОЛЖЭЭ

2023 оны 11-р сарын 19-ны өдрийн 18 цаг 47 минутад Монгол улсын ерөнхийлөгч У.Хүрэлсүхийн дуу хоолойг хиймэл оюун ухаанаар өөрчилж хоолойны өнгө биеийн хэлэмж зэрэгт тааруулж, Хятад хэлээр мэндчилж байгаа хэлбэрт оруулан түгээв. Тус контентийг Өвөрмонголын Талын дуу хоолой радиогийн нийгмийн сүлжээний хуудсаар олон нийтийн хүртээл болгосон бөгөөд 97 минутын хугацаанд нийгмийн сүлжээнд байршаад устгагдсан ч зарим мэдээллийн сайт уг контентийг нийгмийн сүлжээгээрээ давтан түгээжээ.

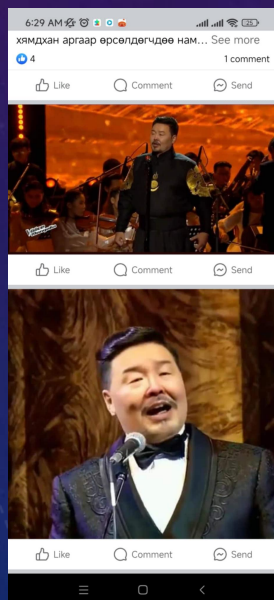
Ингэснээр богинохон хугацаанд уг бичлэг олон мянган хүнд хүрч 983 сэтгэгдэл авчээ. Тус бичлэгийг 2023 оны 09 дүгээр сарын 1-нд ерөнхийлөгч У.Хүрэлсүхийн заасан “Дэлхийн нийслэл Хархорум” сэдэвт теле хичээлийн бичлэгийг ашиглан Hey Gen гэх программыг ашиглан бэлтгэсэн байв.

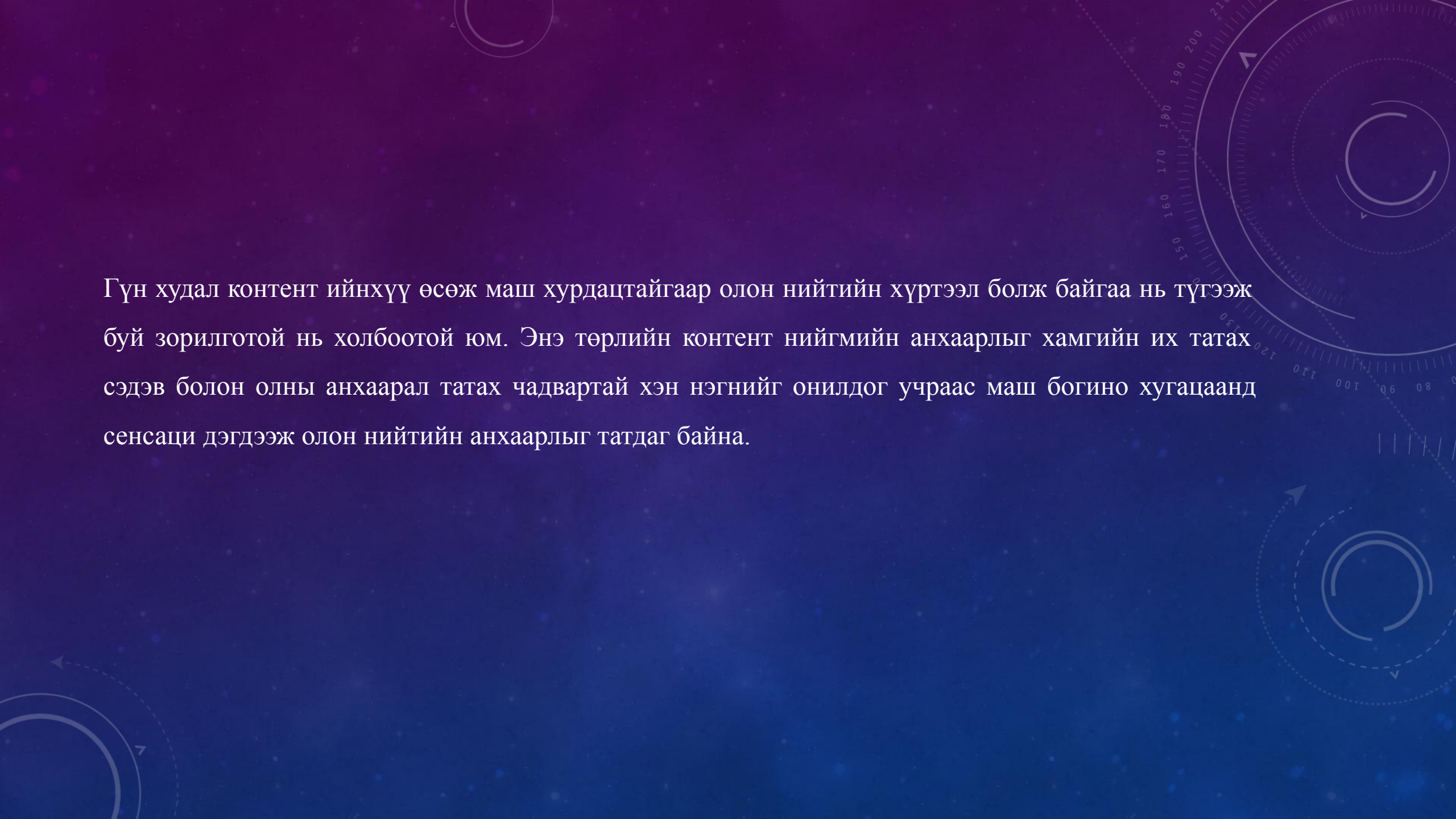
- Тархацын онолын $A + B = AB^2$ гэсэн загварчилалаар тооцоолж үзвэл энэ төрлийн мэдээллийн сүлжээгээр тарах хурд болон цаашид хичнээн хэрэглэгчид тарах боломжийг судалж үзлээ.
- У.Хүрэлсүхийн яриаг өөрчилсөн контентийн давтан үзэгдсэн тооцоолол $983 \text{ (Сэтгэгдэл)} + 576 \text{ (эможи)} = 2430481$ удаа давтан үзэх магадлалтай байна.



- Гүн худал контентийг бэлтгэх технологи манай улсад нэвтрээд сонгуулийн үеэр цөөнгүй асуудал дагуулав. Гэвч худал контентийг зохицуулах хууль эрх зүйн орчин, алготритм, программ хангамж, сэтгүүлчдийн энэ төрлийн худал агуулгатай харьцах, нягтлах дадлага туршлага, иргэдийн хэвлэл мэдээллийн боловсролын түвшин зэрэг манай улсад хангалтгүй
- Иймээс судлаачийн дүгнэлт ёсоор сонгуулийн үеэр энэ төрлийн мэдээлэл газар авах, түгээмэл хэрэглэгдэх явдал нь цаг хугацааны асуудал юм.

Diakopoulos, N. J. (2020). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, 23(7), 1-27. doi:<https://doi.org/10.1177/1461444820925811>.)

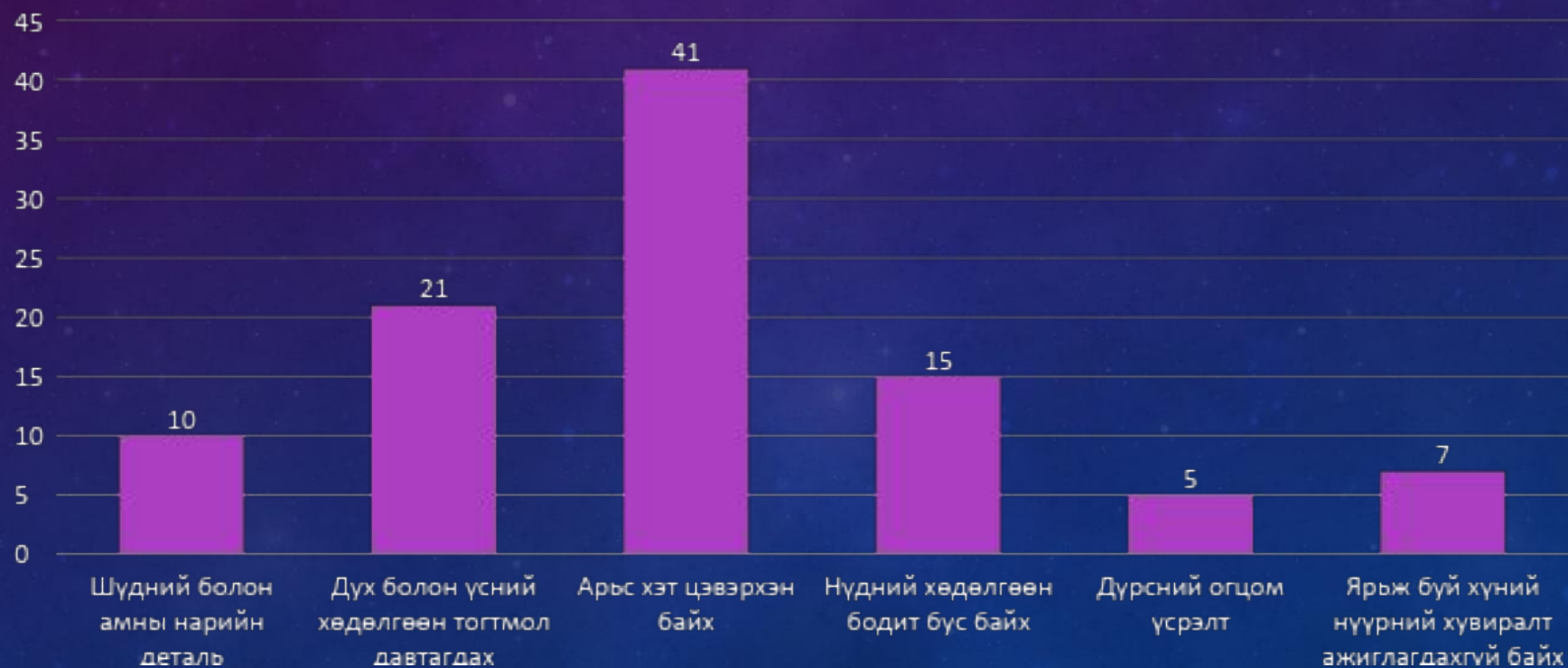




Гүн худал контент ийнхүү өсөж маш хурдацтайгаар олон нийтийн хүртээл болж байгаа нь түгээж буй зорилготой нь холбоотой юм. Энэ төрлийн контент нийгмийн анхаарлыг хамгийн их татах сэдэв болон олны анхаарал татах чадвартай хэн нэгнийг онилдог учраас маш богино хугацаанд сенсаци дэгдээж олон нийтийн анхаарлыг татдаг байна.

ГҮН ХУДАЛ КОНТЕНТООС СЭРГИЙЛЭХ АРГА ЗАМУУД, ТЭДГЭЭРИЙН ТУХАЙ

Бид уг илтгэлийнхээ хүрээнд 14 хоногийн дотор Interface веб хуудсаар түгээгдсэн 6736 гүн худал контентд ажиглалт хийж үзвэл дараах хэдэн төрлийн алдаа ажиглагдаж байв.



Судалгаа хийсэн гүн худал контентийн 67 хувьд нь шүдний болон амны нарийн деталь ажиглагдахгүй байх, дух болон үсний хөдөлгөөн тогтмол давтагдах, арьс хэт цэвэрхэн байх, нүдний хөдөлгөөн бодит бус байх, ярьж буй хүний нүүрний хувирал ажиглагдахгүй байх, дүрсний огцом үсрэлт гэсэн алдаанууд ажиглагдаж байсан бөгөөд эдгээрийг ажиглалт болон алгоритмаар илрүүлэх боломжтой байна. Харин үлдсэн 21 хувьд техниктэй холбоотой алдаа огт ажиглагдаагүй бөгөөд гүн худал контент болохыг зөвхөн логик, дэс дараалал, үйл явдлын өрнөлийг ажигласны үндсэн дээр илрүүлэх боломжтой бөгөөд эдгээрт алгоритмнууд 100 үнэлгээг өгч чадахгүй байна. Үлдсэн 12 хувийг энгийн нүдээр ялгах боломжгүй бөгөөд зөвхөн тусгай программ ашиглан илрүүлж болох гүн худал контентууд байв.

- Гүн худал контентоос сэргийлэхийн тулд нэгдүгээрт алгоритмын түвшинд гүн худал контентоос сэргийлэх арга замыг тооцоолох, хоёрдугаарт сэтгүүлч хэвлэл мэдээллийн салбарын ажилтнууд баримтыг нягтлахдаа үйл явдал, дүрслэлийн логик дэс дарааллыг тогтоох арга зүйг эзэмших гуравдугаарт иргэд хүлээн авагчдын хэвлэл мэдээллийн боловсролыг дээшлүүлэх нь шаардлага тулгарч байна.

АЛГОРИТМЫН ТҮВШИНД ГҮН ХУДАЛ КОНТЕНТООС СЭРГИЙЛЭХ НЬ

- Бид хуурамч контент боловсруулах үе шатыг тодорхойлох үүднээс Монголд болон олон улсын түвшинд нийтлэгдсэн нийт 197 гүн худал контентод судалгаа хийсний үндсэн дээр гүн худал контентыг бий болгох үе шат дараах математик тэгшитгэлийн дагуу бэлтгэгдэж байгааг ажиглалаа. Ингэхдээ бид GeoGebra, Maxima зэрэг программууд ашиглахын зэрэгцээ Принстоны их сургуулийн математикийн тэхнхимийн профессор Shou-Wu Zhang- ын GANs математик загварчилалын тэгшитгэлд ажилд хийж үндсэн дээр бид дараах загварчилалын боловсрууллаа.

Гүн худал мэдээллийг бэлтгэх алгоритмын математик загварчилал

$$L(\mu G, \mu D) := E_{x \sim y \sim \mu D}(x) [\ln y] + E_{x \sim y \sim \mu D}(x) [\ln(1 - y)].$$

$L(\mu G, \mu D)$ – Гүн худал мэдээлэл

μG – Бодит өгөгдөл

μD – Худал өгөгдөл

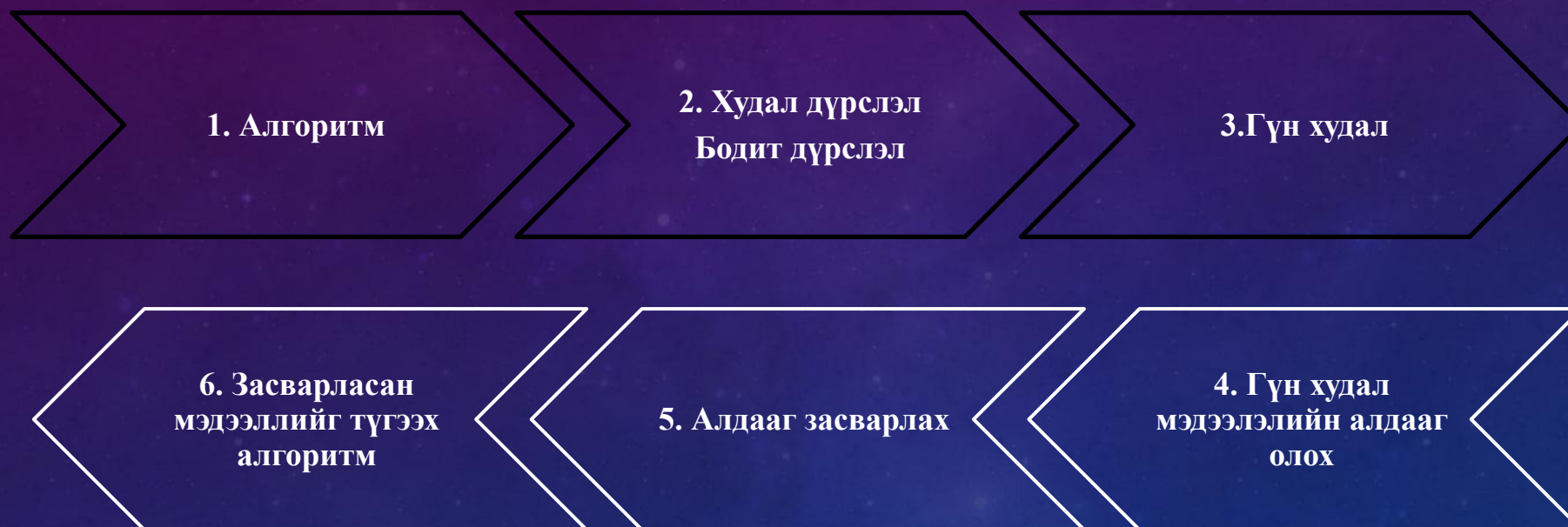
ЗУРАГ №1 МАТЕМАТИК ЗАГВАРЧЛАЛ /ТЭГШИТГЭЛ/ -Д ТУЛГУУРЛААД ГҮН ХУДАЛ КОНТЕНТ БИЙ БОЛОХ ҮЙЛ ЯВЦЫН ЗУРАГЛАЛЫГ БОЛОВСРУУЛСАН БАЙДАЛ



- Дээрх загвараас харвал гүн худал контент бэлтгэдэг алгоритм нь тодорхой мэдээ мэдээлэл бэлтгэх, нийтлэлийн анализ хийх, нийтлэл нэвтрүүлэг бэлтгэх зэрэг модульчилсан үйлдэл хийхдээ асар том хайлтын системтэй адил хэлбэрээр ажиллаад тоглоомын онолын математик тооцоололууд болох ICCT (In computational complexity theory) болон NP (non deterministic polynomial time) математик магадлалын загваруудыг ашиглан ажиллаж, шат дамжлага бүхий тооцооллын үндсэн дээр эцсийн бүтээгдэхүүнийг бэлтгэдэг болох нь харагдаж байна.

БИД ДЭЭРХ СХЕМЧИМЛСАН ЗАГВАР МАТЕМАТИК ТЭГШИТГЭЛД ТУЛГУУРЛААД ГҮН ХУДАЛ КОНТЕНТИЙН ӨӨРӨӨ СУРАЛЦАХ ЧАДВАРЫГ ДАРААХ БАЙДЛААР ТОДОРХОЙЛЖ ҮЗЛЭЭ.

Зураг №2 Гүн худал контентийг илрүүлэх программ хангамжийн загварчилал



Жишээлбэл, 2018 онд гарсан Доналд Трампын порнограф бичлэг 2024 онд яг тэр загвараараа хийгдсэн ч бодит байдлын тал дээр өөрчлөгдөж худал гэдэг нь мэдэгдэхээргүй болжээ.



- Бид тус илтгэлд алгоритмын тооцооллыг боловсрууллаа. Иймээс үүний дагуу манай улс худал контент бэлтгэдэг алгоритмын ажиллагаа, математик тэгшитгэлд суурилсан гүн худал контентийг таних алгоритм бүхий программ хангамжуудыг боловсруулах, тэдгээрийг иргэд болон сэтгүүлч редакторуудын хэрэглээнд нэвтрүүлэх шаардлагатай байна.
- Эсвэл үндэсний хэмжээний нууцын тухай хуульд хамаарахгүй өгөгдлийн санг олон улсад үйл ажиллагаа явуулдаг компаниудад бүтэцчилэгдсэн хэлбэрээр боловсруулан өгч гүн худал контенттой тэмцэхэд нь хамтарч ажиллавал уг асуудлаас сэргийлэх боломжтой юм.

СЭТГҮҮЛЧИД МЭРГЭЖЛИЙН УР ЧАДВАРАА ХӨГЖҮҮЛЖ, МЭДЭЭЛЛИЙГ МЭРГЭЖЛИЙН ТҮВШИНД БЭЛТГЭХ ЗАМААР ГҮН ХУДАЛ АГУУЛГААС СЭРГИЙЛЭХ НЬ

- Гүн худал контент ихэвчлэн нийгмийн сүлжээгээр тардаг. Гэвч энэ төрлийн агуулга маш хурдацтайгаар олон нийтийн хүртээл болоход хэвлэл мэдээллийн хэрэгслүүд сэтгүүлчид үүрэг гүйцэтгэсээр байна. Бидний судалгаагаар иргэдийн түгээсэн худал контентоос сэтгүүлч хэвлэл мэдээллийн хэрэгслийн түгээсэн худал контентийн үзэлт, тархалт 4-6 дахин өндөр үзүүлэттэй байгаа нь ажиглагдав.
- Уг нь гүн худал контент нь огт болоогүй үйл явдал байдаг. Иймээс өдөр тутмын үйл явдалд оролцож тогтмол мэдээ, мэдээлэл сурвалжилдаг, тухайн салбараар мэргэшсэн сэтгүүлчид үйл явдлыг логик дэс дараалалд нь дүгнэлт хийсний үндсэн дээр энэ төрлийн контентийг ялгаж таних боломж бий. Мөн Олон улсын дата сэтгүүл зүйн төвийн боловсруулсан жишгийн дагуу сэжигтэй гэж үзсэн мэдээлэлд анализ хийж баримтыг нягтлах нь чухал юм.

- Хуурамч мэдээлэл нь хүнийг хуурах зориулалттай түр олны анхаарлыг татах агуулга байдаг. Иймээс энэ төрлийн контент бэлтгэж байгаа этгээдүүд нь ихэвчлэн мэргэжлийн сэтгүүлчид биш юм. Иймээс мэдээллийн агуулга бэлтгэсэн хэлбэр хийц нь сэтгүүл зүйн нийтлэл нэвтрүүлэгт тавигддаг үндсэн шалгуур шаардлагыг хангах тохиолдол харьцангуй бага. Иймээс сэтгүүлчид нэг талаар мэдээлэлд мэргэжлийн үүднээс үнэлэлт дүгнэлт өгч, баримт мэдээллийг мэргэжлийн үүднээс нягтлах нь гүн худал контентийг түгээхээс сэргийлэх нөгөө талаар сэтгүүлчид мэдээллийг сэтгүүл зүйн шалгуур шаардлагын үүднээс бэлтгэх нь гүн худал контентоос сэргийлэх үндсэн арга болж байна.

ИРГЭДЭД ХЭВЛЭЛ МЭДЭЭЛЛИЙН БОЛОВСРОЛ ОЛГОХ ЗАМААР ГҮН ХУДАЛ КОНТЕНТООС СЭРГИЙЛЭХ НЬ

Гүн хуурамч контентийг алгоритмын түвшинд хянаад, сэтгүүлчид мэргэжилтнүүд ялган таньж мэдээллийн хэрэгслүүд болон нийгмийн сүлжээнээс устгасан ч иргэд уг агуулгад итгэх, санаатай болон санамсаргүйгээр дахин түгээх, явган хов болгох, үйлдвэрлэж нийтийн хүртээл болгох ажиллагааг зогсоохгүй бол асуудлыг шийдэхэд бэрхшээлтэй.

- Иргэдийн тараасан гүн хуурамч контент хэдийгээр бүх нийтийг хамарч чадахгүй ч тодорхой хүрээлэлд түгэх замаар удаан боловч аажмаар түгдэг. Мөн хэвлэл мэдээллийн хэрэгслүүд болон хуурамч аккоунтуудын нийгмийн сүлжээнд түгээсэн худал контентыг хадгалж аваад цахим орчноос устаад олон нийтийн анхаарлаас гарсны дараа дараа дахин түгээх явдал ч байсаар байна. Тухайлбал, бидний сонгож авсан хэд хэдэн гүн хуурамч контент албан ёсны гэх эх сурвалжуудаас утсан ч хувийн аккоунтуудаар дамжин нийтийн хүртээл болсоор байгаа нь ажиглагдлаа.

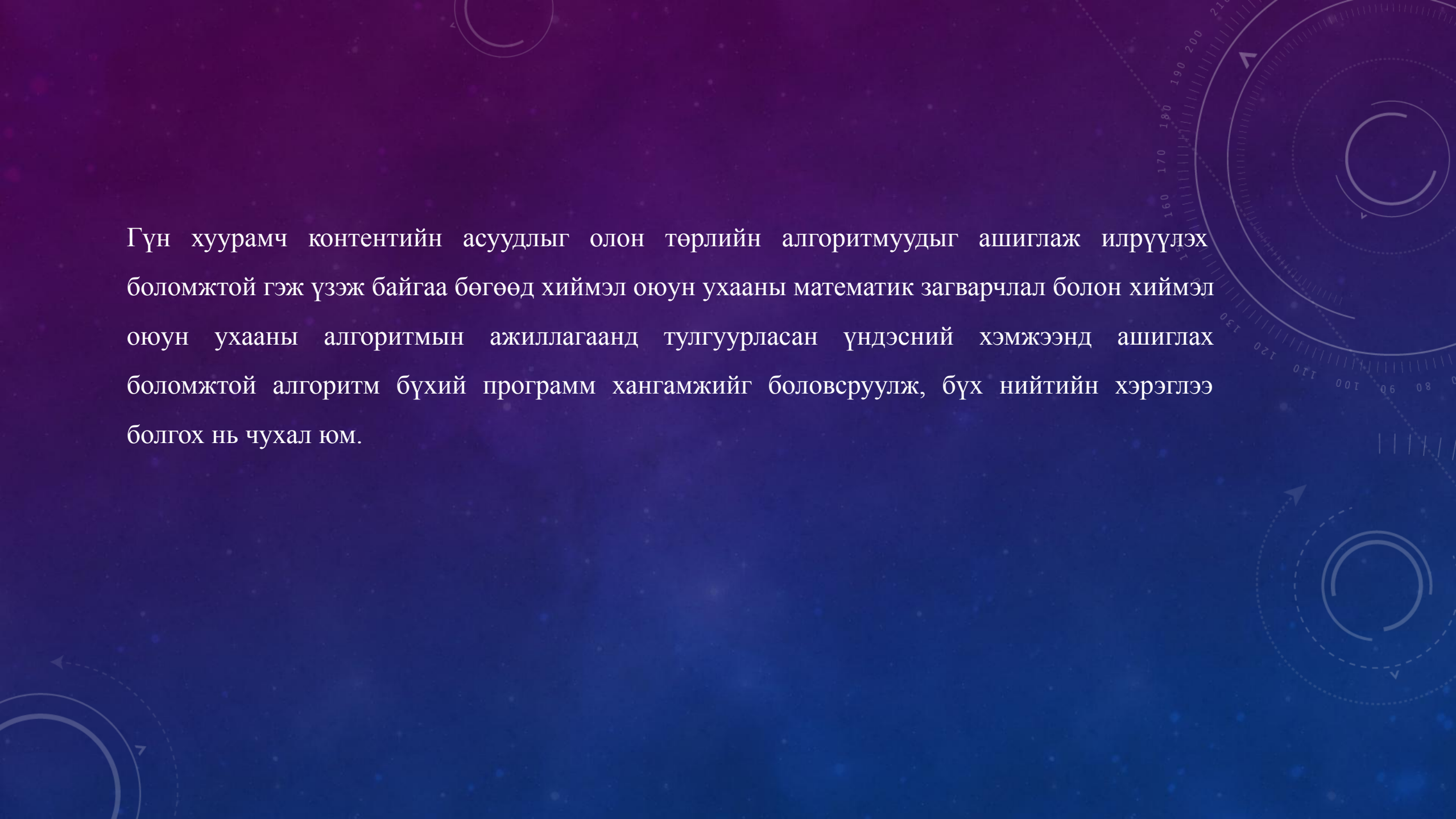
- Монгол Улсын ерөнхийлөгч У.Хүрэлсүхийн Хятадаар ярьсан гэх бичлэг нийгмийн сүлжээнд устгагдсанаасаа хойш 4-5 удаа дахин нийтлэгдсэн бөгөөд хэдийгээр түгээлт тус бүр дээр үзэлт сэтгэгдэл нь буурч байгаа ч үзэлтийн тоо 100-250 мянган хооронд хэлбэлзэж байгаа нь харагдав. Ялангуяа Facebook сүлжээнд Reels хэлбэрээр байрлуулсан хувилбар нь 246 мянган үзэлт, мянга орчим сэтгэгдэлтэй байгаа нь ажиглагдлаа. Гэхдээ тус мэдээллийг мэдээллийн хэрэгсэл биш 500 хүрэхгүй дагагчтай хувийн аккоунтаас түгээхэд үзэлт ийм өндөр байсаар байна.

- Visual Social, Loomly зэрэг программуудыг ашиглан У.Хүрэлсүхтэй холбоотой Reels -ийн сэтгэгдэлд ажиглалт хийж үзвэл хэдийгээр устгагдаад дахин түгээгдэж байгаа боловч 978 сэтгэгдлийн 343 нь сэжигтэй буюу хуурамч байж болох аккоунтаас бичигдсэн байгаа ч үлдсэн 635 сэтгэгдлийг бодит байж болох аккоунтаас нийтэлжээ.
- Тус контетд нийтлэгдсэн бүх сэтгэгдлийг SPSS программд оруулж эерэг (худал эсэхийг мэдсэн) сөрөг (худал мэдээлэлд итгэсэн дэмжсэн) гэсэн томъёололд оруулж тооцож үзвэл бодит байж болох аккоунтын бичсэн нийт сэтгэгдлийн 93 хувь буюу 594 нь сөрөг агуулга бүхий сэтгэгдэл байлаа.
- Худал болохыг нь тогтоогоод хэвлэл мэдээллийн хэрэгслүүдийн нийтэлсэн мэдээлэлд сөрөг агуулга бүхий буюу мэдээллийг үнэн гэж итгэсэн агуулга бүхий сэтгэгдэл цөөнгүй байна.

ДҮГНЭЛТ

Хиймэл оюун ухаан нь сэтгүүл зүйн салбарт олон шинэчлэлийг авчирлаа. Ялангуяа машин сургалтын аргад үндэслэсэн бий болсон бичил хиймэл оюун ухааны ололтууд сэтгүүлчдийн ажлыг хөнгөвчлөх арга хэрэгслүүд болж, өдөр тутмын ажилд туслагчийн үүргийг гүйцэтгэж байна. Гэхдээ хиймэл оюун ухаан нь хөгжил дэвшлээс гадна хэвлэл мэдээллийн салбарт худал зохиомол мэдээллийн цоо шинэ төрөл болох гүн худал контентийг бий болгожээ. Энэ төрлийн контент нь бие мах бодийг өөрчилсөн, үг яриаг өөрчилсөн хэлбэрээр жендэр болон гарал үүсэл, үндэстэн ястнаар ялгаварлан гадуурхах, порнограф, нийгмийн халамж зэрэг агуулгатайгаар мэдээллийн шинэ хэрэгслүүд болон нийгмийн сүлжээгээр түгэж байна. Гүн худал контент нь бодит мэдээллээс ялгагдахааргүй адилхан түвшинд хийгдэх болсон нь үнэн худлын зааг алга ялгааг алга болгож, мэдээллийн үнэ цэн хэвлэл мэдээллийн хэрэгслийн нэр төрд сөргөөр нөлөөлөх болов.

Бие мах бодийг өөрчилсөн энэ гүн худал контент нь порнограф хэлбэртэй байх тохиолдол элбэг байна. Мөн зүс царай бие мах бодийг нь шоглосон эмэгтэйчүүд рүү чиглэсэн агуулга зонхилж байгаа нь харагдав. Үг яриаг өөрчилсөн контентын хувьд улс төрчдийг шоглосон гарал үүсэл, хувь хүний үзэл баримтлалтай холбоотой агуулгыг илэрхийлсэн, илт доромжилсон олны өмнө алдар хүндийг нь гутаах зорилготой байна.

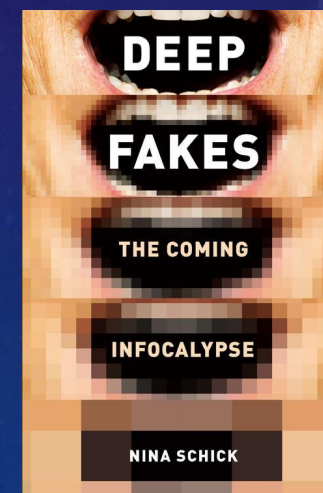


Гүн хуурамч контентийн асуудлыг олон төрлийн алгоритмуудыг ашиглаж илрүүлэх боломжтой гэж үзэж байгаа бөгөөд хиймэл оюун ухааны математик загварчлал болон хиймэл оюун ухааны алгоритмын ажиллагаанд тулгуурласан үндэсний хэмжээнд ашиглах боломжтой алгоритм бүхий программ хангамжийг боловсруулж, бүх нийтийн хэрэглээ болгох нь чухал юм.

Хэдийгээр гүн хуурамч контент нь нийцмийн сүлжээгээр тархдагч хэвлэл мэдээллийн хэрэгслээр дамжин олны хүртээл болох явдал цөөнгүй байна. Иймээс бодит ба хуурамч контентийн хооронд тодорхой хил хязгаар тогтоохын тулд сэтгүүлчид алгоритмд суурилсан шинэ арга зүйгээс гадна үйл явдлын дэс дараалал логик уялдаа холбоог тогтоох замаар хуурамч контентийг илрүүлэх арга зүйд суралцаснаар сэргийлэх боломжтой.

Иргэд гүн хуурамч контентийг дахин түгээх, сэтгэгдэл бичих, эможи илэрхийлэх хэлбэрээр тархалтад нөлөө үзүүлсээр байгаа нь судалгаанаас харагдлаа. Иймээс иргэдэд хэвлэл мэдээллийн боловсрол олгож “хортой” мэдээллээс өөрсдийгөө хамгаалах арга зүйд сургах, хэвлэл мэдээллийн боловсрол олгох агуулга бүхий модуль хичээлийг хичээлийг ЕБС, насан туршийн боловсролын төвүүд, коллеж, их дээд сургуулийн сургалтын хөтөлбөрт тусгах хэрэгтэй байна.

- Эцэст нь тэмдэглэвэл гүн хуурамч контенттой яг одооноос эхэлж тэмцэхгүй бол иргэдийн мэдээ мэдээлэлд итгэх нэг мөсөн буурах цаашлаад мэдээллийн сүйрэл (Infocalypse) бий болоход ч ойрхон байна. Ялангуяа манай улсын иргэд хэвлэл мэдээллийн боловсролоор ихээхэн дутмаг байгаа талаар олон судалгаанд тэмдэглэгдсэн байдаг. Ийм учраас энэ төрлийн мэдээлэл манайд газар авбал сэргийлэх болон хянахад бэрхшээл үүсэх нь ойлгомж юм.

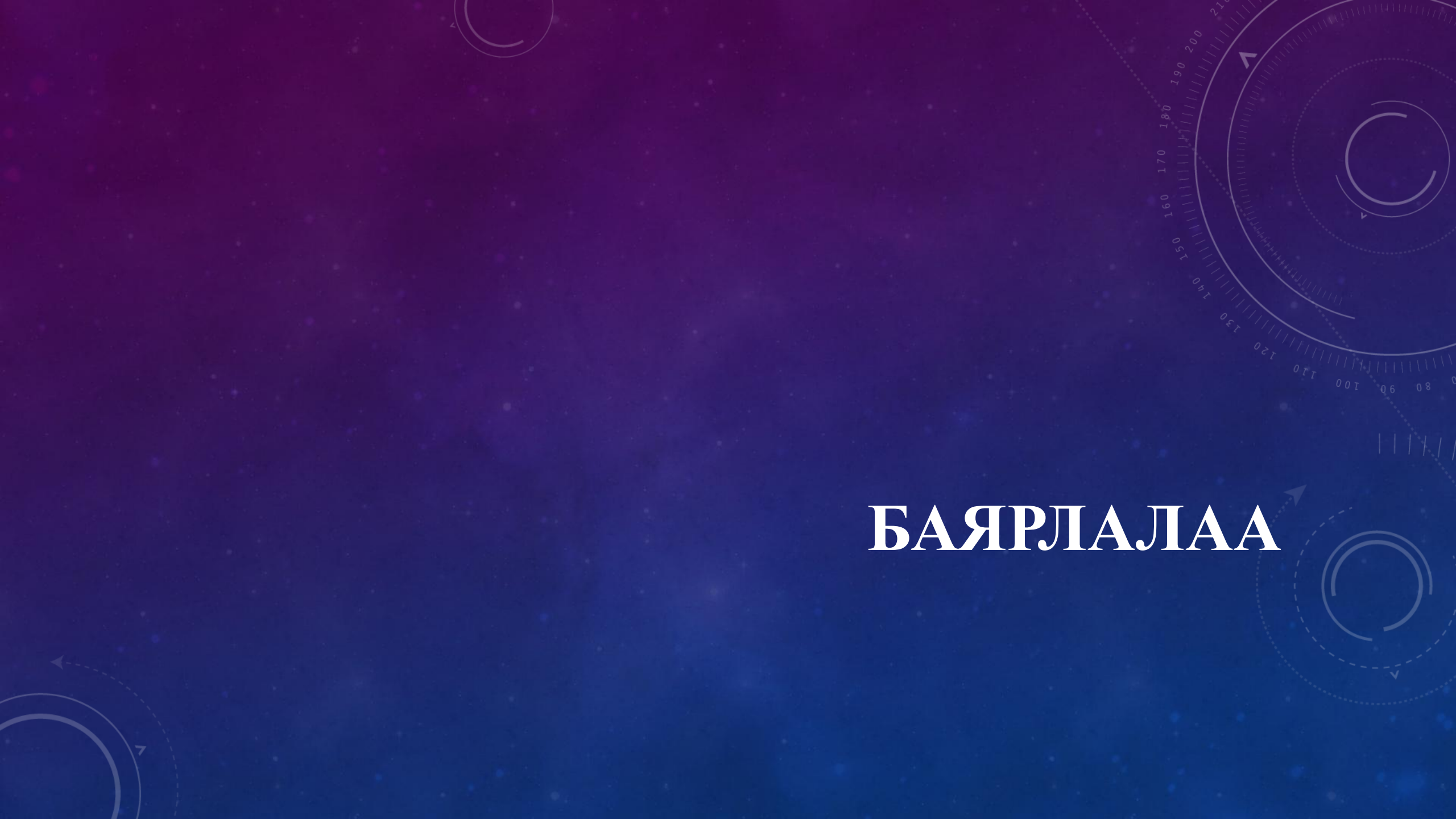


НОМ ЗҮЙ

- A.Boden, M. (2019). Artificial intelligent: A very short introduction. Oxford: Oxford university press.
- Binmore, K. (2007). Theory of game. Oxford: Oxford university press.
- Carmel Kent, B. d. (2022). AI for Learning. London: CRC press.
- Diakopoulos, N. J. (2020). Anticipating and addressing the ethical implications of deepfakes in the context of elections. New Media & Society, 23(7), 1-27. doi:<https://doi.org/10.1177/1461444820925811>.)
- Don Tapscott, A. T. (2018). Blockchain Revolution. New York: An inprint Penguin Random House LLC.
- E.Holms, D. (2017). Big data: Very short introduction. Oxford: Oxford University press.

- European Journalism Center. (2023). Verification Handbook For Disinformation And Media Manipulation. EU: Data-Journalism.com.
- Gosse, C. B. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37 (5), 497-511. doi: <https://doi.org/10.1080/15295036.2020.1832697>
- J.Hand, D. (2008). *Statistics*. Oxford: Oxford university press.
- Priest, G. (2017). *Logik*. Oxford: Oxford university press.
- Riparbelli, V. (2023 оны 11 1). *Synthesia*. The Future of (Synthetic) Media: <https://www.synthesia.io/post/the-future-of-synthetic-media>

- Schwab, K. (2017). The Fourth Industrial Revolution . Davos: World economic forum.
- Simpson, B. N. (2023). The tensions of deepfakes, Information,. Communication & Society. doi: DOI: 10.1080/1369118X.2023.2234980
- Van Der Nagel, E. (2020). Verifying images: Deepfakes, control, and consent. Porn Studies, (4), 424-429. doi:<https://doi.org/10.1080/23268743.2020.1741434>.)
- Wakefield, J. (2022 оны 3 18). Deepfake presidents used in Russia-Ukraine war. BBC. <https://www.bbc.co.uk/news/technology-60780142>
- Yadlin-Segal, A. O. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. Convergence: The International Journal of Research Into New Media Technologies, 27(1), 36-51. doi:<https://doi.org/10.1177/1354856520923963>)



БАЯРЛАЛАА