



# ХИЙМЭЛ ОЮУН УХААНЫ ЗАСАГЛАЛЫН ТОГТОЛЦООНЫ ЗАГВАРЧЛАЛ

**SG:D**  
EMPOWERING POSSIBILITIES

**IM** INFOCOMM  
MEDIA  
DEVELOPMENT  
AUTHORITY

**pdpc** PERSONAL DATA  
PROTECTION COMMISSION  
SINGAPORE



## АГУУЛГА

---

<b>ХУРААНГУЙ</b> .....	4
<b>ОРШИЛ ҮГ</b> .....	7
<b>1. УДИРТГАЛ</b> .....	9
<b>2. ТАНИЛЦУУЛГА</b> .....	12
Зорилт.....	13
Загварчлалын удирдамж, зарчим.....	15
Таамаглал.....	17
Тодорхойлолт.....	18
<b>3. ХИЙМЭЛ ОЮУН УХААНЫ ЗАСАГЛАЛЫН ТОГТОЛЦООНЫ ЗАГВАРЧЛАЛ</b> .....	19
Дотоод засаглалын бүтэц, хэмжүүр.....	21
Хиймэл оюун ухааны шийдвэр гаргалтанд хүний оролцооны түвшинг тодорхойлох.....	28
Үйл ажиллагааны менежмент.....	35
Оролцогч талуудын хамтын ажиллагаа.....	53
<b>ХАВСРАЛТ “А”</b>	
Зөвлөмж: Хиймэл оюун ухааны ёс зүйн зарчмуудын эмхэтгэл.....	64
<b>ХАВСРАЛТ “Б”</b>	
Алгоритм аудит /үнэлэх, хянах/.....	67
<b>ТАЛАРХАЛ</b> .....	68

# Шинэчилсэн ЭМХЭТГЭЛ

НИЙТЛЭЛ	НИЙТЛЭГДСЭН ОГНОО	ТОВЧ УТГА
Эхний хэвлэл	23 Нэгдүгээр сар 2019	2019 онд Швейцарийн Давос хотноо зохион байгуулагдсан Дэлхийн эдийн засгийн чуулга уулзалтаар хиймэл оюун ухааны удирдлагын тогтолцооны загварчлалыг (Эхний хэвлэл) баталсан.
Хоёр дахь хэвлэл	21 Нэгдүгээр сар 2020	2020 онд Швейцарийн Давос хотноо зохион байгуулагдсан Дэлхийн эдийн засгийн чуулга уулзалтаар хиймэл оюун ухааны удирдлагын тогтолцооны загварчлалыг (Хоёр дахь хэвлэл) баталсан.

Голлох үндсэн өөрчлөлтүүд :

- Байгууллагууд хиймэл оюун ухааны удирдлага, засаглалын туршлагыг практик хувилбаруудад хэрхэн хэрэгжүүлснийг харуулахын тулд салбар тус бүрийн жишээг оруулав.
- Хоёр бүлгийн нэрсийг агуулгын хувьд өөрчилж, нарийвчлан тусгах:
  - » "Хиймэл оюун ухаанаар шийдвэр гаргах загварыг тодорхойлох нь" хэсгийг "Хиймэл оюун ухааны шийдвэр гаргалтанд хүний оролцооны түвшинг тодорхойлох" гэж өөрчлөх.
  - » "Хэрэглэгчийн харилцааны менежмент" хэсгийг "Оролцогч талуудын оролцоо ба хамтын ажиллагаа" болгон өөрчлөх.

Дараах хэсгүүдэд тодорхой өөрчлөлтүүдийг оруулсан.

Үүнд:

## **Хиймэл оюун ухааны шийдвэр гаргалтанд хүний оролцооны түвшинг тодорхойлох**

- Хиймэл оюун ухаанаар сайжруулсан шийдвэр гаргахад хүний хяналт “Хүн ба машин хоорондын тасралтгүй харилцан үйлчлэл” гэсэн ойлголтыг тодорхойлсон.
- Байгууллагууд хиймэл оюун ухаантай холбоотой шийдвэр гаргах үйл явцад хүний оролцооны түвшинг тодорхойлохдоо хохирлын мөн чанар, буцаах чадвар, үйл ажиллагааны боломж зэрэг бусад хүчин зүйлсийг харгалзан үзэж болохыг тодорхойлсон.



НИЙТЛЭЛ	НИЙТЛЭГДСЭН ОГНОО	ТОВЧ УТГА
ХОЁР ДАХЬ ХЭВЛЭЛ	21 Нэгдүгээр сар 2020	<p><b>Үйл ажиллагааны менежмент</b></p> <ul style="list-style-type: none"> <li>• Байгууллагуудад дараах арга хэмжээг хэрэгжүүлэхдээ эрсдэлд суурилсан арга барилыг хэрэгжүүлэх арга зам, зааварчилгааг өгсөн. <ul style="list-style-type: none"> <li>» Оролцогч талуудад хамгийн их нөлөө үзүүлэх онцлог, функцуудыг тодорхойлох;</li> <li>» Оролцогч талуудтай итгэлцлийг бий болгоход ямар арга хэмжээ хамгийн үр дүнтэй болохыг авч үзэх.</li> </ul> </li> <li>• Хиймэл оюун ухааны загваруудыг бий болгох хүрээнд эдгээр загваруудын үр ашиг, найдвартай байдал, ёс зүйн нийцлийг хангахын тулд янз бүрийн арга хэмжээний хэрэгцээ шаардлага, ач холбогдлын талаарх зөвлөмжийг тусгасан: <ul style="list-style-type: none"> <li>» Хиймэл оюун ухааны загварыг бүтээхэд ашигладаг өгөгдлийн багцад хувийн болон хувийн бус өгөгдлийг хамтад нь хэрэглэхийг тодорхойлсон.</li> <li>» Бат бөх байдал, дахин давтагдах чадвар, аудит хийх зэрэг шинэ арга хэмжээг тусгахдаа эдгээр талыг үр дүнтэй хэрэгжүүлэхийн тулд тодорхой удирдамж, практикийг бий болгох, хэрэгтэй туршлагауудын жишээг оруулж өгсөн.</li> </ul> </li> </ul> <p><b>Оролцогч талуудын итгэлцэл, хамтын ажиллагаа</b></p> <ul style="list-style-type: none"> <li>• Дотоод болон гадаад талуудтай хамтран ажиллах ач холбогдлыг онцлон тэмдэглэв.</li> <li>• Оролцогч талуудтай харилцахдаа зорилго, нөхцөл байдлыг харгалзан үзэх шаардлагатайг онцлон тэмдэглэв.</li> <li>• Оролцогч талуудтай харилцахдаа ямар түвшинд хүргэх мэдээллийн талаарх санал зөвлөмж.</li> </ul> <p><b>Хавсралт А – Баримт бичгийн лавлах: Хиймэл оюун ухааны ёс зүйн зарчмуудын эмхэтгэл</b></p> <ul style="list-style-type: none"> <li>• <u>Хавсралт А</u> Хиймэл оюун ухааны ёс зүйн журмын эмхэтгэл бөгөөд энэхүү хавсралтаар хиймэл оюун ухааны үндсэн загварт тусгагдсан журмуудаас гадна өөрийн байгууллагын дотоод журамд нэмэлт өөрчлөлт оруулж тусгах боломжтой баримт бичиг юм.</li> </ul>

---

НИЙТЛЭЛ	НИЙТЛЭГДСЭН ОГНОО	ТОВЧ УТГА
---------	----------------------	-----------

---

Хоёр дахь  
хэвлэл

21 Нэгдүгээр сар  
2020

**Хавсралт Б – Алгоритмын аудит**

- Алгоритмын бодит үйлдлийг илрүүлэх зайлшгүй шаардлагатай тохиолдолд алгоритмын аудитыг зөвхөн зохицуулагчийн хүсэлтээр, эрх бүхий байгууллагын хяналтан доор хийх ёстойг удирдамжид тодорхойлсон.

**Хавсралт С – Хэрэглээний тохиолдол**

- Хавсралт С хасагдсан. Үүний оронд тусад нь хэрэглээний тохиолдлуудын эмхэтгэл нийтлэгдсэн. [/go.gov.sg/ai-gov-use-cases/](https://go.gov.sg/ai-gov-use-cases/)
-

## ӨМНӨХ ҮГ

2019 онд дэлхий даяар хиймэл оюун ухааныг (AI) боловсронгуй болгож, өргөнөөр ашиглахад мэдэгдэхүйц ахиц дэвшил гарсан жил болж байна. Онцлох жишээ нь хиймэл оюун ухаан хурдацтай хөгжиж, өдөр тутмын амьдралын янз бүрийн талбарт өсөн нэмэгдэж буй интеграцчлалыг онцолж байна. Бид GPT-2 гэх мэт хиймэл оюун ухаанаар ажилладаг шинэ үеийн байгалийн текст үүсгэгчүүд гарч ирсний гэрч болсон бөгөөд тэдгээр нь ихэвчлэн ялгагдахгүй уялдаатай, контекст хамааралтай текстийг бүтээх гайхалтай чадварыг харуулсан. Мөн OpenAI-ийн бүтээсэн Dactyl хэмээх робот гарыг хөгжүүлснээр робот техник, бататгах сургалтад томоохон үсрэлт хийсэн. Dactyl энгийн гэр ахуйн эд зүйлсийг хүний гартай дүйцэхүйц хэмжээний ур чадвараар барьж, удирдах чадварыг харуулсан. Эдгээр жишээнүүд нь хиймэл оюун ухаан хурдацтай хөгжиж, бидний өдөр тутмын амьдралд улам бүр түгээмэл болж байгааг онцолж байна.

Хиймэл оюун ухааны ёс зүй ба засаглалын талаарх яриа сүүлийн жилүүдэд үнэхээр ахисан. Засгийн газар болон олон улсын байгууллагууд хиймэл оюун ухааны ёс зүйн үр дагавар, засаглалын хэрэгцээг шийдвэрлэхийн тулд зарчим, тогтолцоо, зөвлөмж гаргахад улам бүр анхаарлаа хандуулж байна. 2019 оны 1-р сард Давост болсон Дэлхийн эдийн засгийн чуулга уулзалтад Сингапур улс хиймэл оюун ухааны засаглалын тогтолцоог танилцуулснаар энэ хүрээнд анхаарал татахуйц өөрчлөлт гарсан. Тогтолцооны загварчлал нь ёс зүйн зарчмуудыг хэрэгжүүлэх боломжтой болгох арга барилаараа ихээхэн онцлогтой. Зөвхөн онолын үзэл баримтлалыг тоймлохын оронд хиймэл оюун ухааныг хариуцлагатай нэвтрүүлэхийн тулд байгууллагууд хялбархан хэрэгжүүлж болох практик зөвлөмжийг санал болгосон. Энэхүү практик чиг баримжаа нь янз бүрийн байгууллагуудад үүнийг нэвтрүүлэхэд чухал үүрэг гүйцэтгэсэн бөгөөд энэ нь хэрэглэгчдэд ээлтэй, өөр өөр салбаруудад хамааралтай болохыг онцолсон юм. Загварын тогтолцоонд тусгагдсан туршлагыг өргөнөөр нэвтрүүлж байгаа нь хиймэл оюун ухааны ёс зүй, засаглалын цогц орчинд удирдахад нь байгууллагуудад туслах үр дүнтэйг харуулж байна.

Сингапур хиймэл оюун ухааны засаглалын загвар тогтолцооны хоёр дахь хэвлэлийг танилцуулж байгаадаа баяртай байна. Энэхүү шинэчлэгдсэн хэвлэлд хиймэл оюун ухаан нэвтрүүлсэн байгууллагуудын үнэ цэнэтэй туршлагыг нэгтгэж, Европын Комиссын Өндөр түвшний шинжээчдийн бүлэг, ЭЗХАХБ-ын хиймэл оюун ухааны шинжээчдийн бүлэг зэрэг олон улсын тэргүүлэх платформуудад оролцсоны үр дүнд олж авсан санал хүсэлтийг тусгасан. Эдгээр хувь нэмэр нь хиймэл оюун ухааныг хариуцлагатай хэрэгжүүлэх байгууллагуудад илүү ойлгомжтой, үр дүнтэй удирдамж санал болгох боломжийг бидэнд олгож байна.

Сингапурын IMDA (**Infocomm Media Development Authority**), PDPC (**Personal Data Protection Commission**) болон Дэлхийн эдийн засгийн форум хоорондын хамтын ажиллагааны үр дүнд Байгууллагуудын хэрэгжилт, өөрийн үнэлгээний гарын авлага (**ISAGO- Implementation and Self-Assessment Guide for Organizations**), мөн Хэрэглээний тохиолдлын эмхэтгэл зэрэг үнэ цэнэтэй нөөцийг бий болгосон. Хиймэл оюун ухааныг хэрэгжүүлэхэд тулгарч буй нарийн төвөгтэй асуудлуудыг хариуцлагатай, үр дүнтэйгээр удирдах боломжийг байгууллагуудад иж бүрэн арга барилаар хангадаг. Эдгээр удирдамжийг дагаж, бодит жишээнээс суралцсанаар байгууллагууд хиймэл оюун ухааны засаглалын хүрээгээ сайжруулж, хиймэл оюун ухааныг ёс зүйтэй, хариуцлагатай ашиглах өргөн хүрээний зорилгод хувь нэмрээ оруулах боломжтой.

Эдгээр санаачилгууд нь Сингапурын хиймэл оюун ухааны үндэсний стратегид чухал үүрэг гүйцэтгэдэг. Олон нийтийн итгэлийг бий болгож, тогтвортой байлгах хиймэл оюун ухааны засаглалд хүн төвтэй хандлагыг хөгжүүлэх бидний төлөвлөгөөний жишээ юм. Цаашилбал, тэд хиймэл оюун ухааны экосистемийг хамтран бий болгоход анхаарлаа хандуулж байгааг онцлон тэмдэглэв. Тогтолцооны загварчлал ба ISAGO нь хиймэл оюун ухааныг ёс суртахууны хэрэглээнд сургах мэргэжлийн боловсон хүчнийг сургах зэрэг цаашдын ахиц дэвшлийн эхлэлийг тавьсан. Эдгээр хүчин чармайлт нь Сингапур, магадгүй дэлхий дахинд хиймэл оюун ухааны нийгэмд үзүүлэх нөлөөллийг илүү үр дүнтэй удирдах боломжийг олгоно.

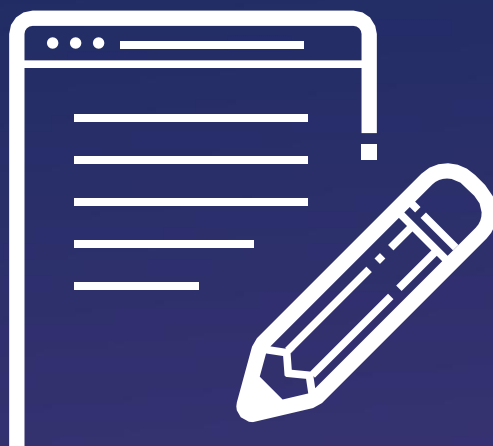
Өнөөдөр бидний хийж буй үйл ажиллагаа нь бидний хамтын ирээдүйг гүн гүнзгий байдлаар тодорхойлох болно. Тогтолцооны загварчлал нь хиймэл оюун ухааныг хариуцлагатай нэвтрүүлэх, тасралтгүй хөгжүүлэх үндсэн тулгуур юм. Энэхүү суурь дээр тулгуурлан бид хиймэл оюун ухаанд хүн төвтэй хандлагыг урагшлуулахаар зорьж байгаа бөгөөд энэ нь олон нийтийн итгэлийг хангах, инновацийг дэмжиж, хиймэл оюун ухааны чадавхийг ирээдүй хойч үедээ ашиглах явдал юм.

## **S Iswaran**

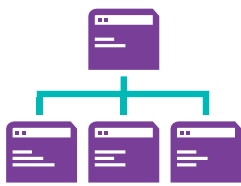
*Minister for Communications and Information Singapore*

*January 2020*





# 1. ОРШИЛ



- 1.1 Тогтолцооны загварчлал нь дотоод засаглалын бүтэц, арга хэмжээ, хиймэл оюун ухааныг сайжруулсан шийдвэр гаргахад засаглал, хүний оролцоо, үйл ажиллагааны удирдлага, оролцогч талуудын оролцоо гэсэн дөрвөн өргөн хүрээнд төвлөрсөн хиймэл оюун ухааны хариуцлагатай хэрэглээг сайжруулахад чиглэсэн цогц гарын авлага юм.

Хэдийгээр Тогтолцооны загварчлал нь хиймэл оюун ухааныг хариуцлагатай ашиглах үнэ цэнэтэй удирдамжийг санал болгодог ч түүний хэрэглээ нь хэд хэдэн практик болон нөхцөл байдлын хязгаарлалттай байдаг. Байгууллагууд эдгээр сэрэмжлүүлгийг хүлээн зөвшөөрч, хиймэл оюун ухааны технологийн динамик шинж чанар, ёс зүйн стандартад нийцүүлэхийн тулд байнгын үнэлгээ, шинэчлэлтийг баталгаажуулж, тогтолцоог өвөрмөц нөхцөл байдалд нь тохируулах ёстой. Үүнийг харгалзан хэд хэдэн анхааруулгыг дурдах нь зүйтэй. Тогтолцооны загварчлал нь –

**a. Алгоритм-агностик**

Тодорхой хиймэл оюун ухаан эсвэл өгөгдлийн аналитик арга зүйд анхаарлаа хандуулдаггүй. Үүний оронд хиймэл оюун ухааны системийг ерөнхийд нь дизайн, хэрэглээ, хэрэглээнд өргөн хүрээнд хамаарна.

**b. Технологи-агностик**

Ямар нэгэн тодорхой систем, программ хангамж, технологиос хамааралгүй олон талт байхаар бүтээгдсэн. Энэ нь янз бүрийн технологийн орчинд өргөн нийцтэй байдал, уян хатан байдлыг хангадаг хөгжүүлэлтийн хэл, өгөгдөл хадгалах аргаас үл хамааран хэрэгжинэ.

**c. Сектор-агностик**

Энэхүү тогтолцоо нь салбараас үл хамааран аливаа байгууллага хэрэгжүүлэх боломжтой анхаарах зүйлс, арга хэмжээний суурь болж өгдөг. Энэ нь тодорхой салбар, байгууллагын өвөрмөц хэрэгцээг хангахын тулд нэмэлт хүчин зүйлүүдийг харгалзан өөрчлөх боломжтой уян хатан суурь үзүүлэлтийг өгдөг. Хувийн Мэдээллийг Хамгаалах Хороо (PDPC) нь төрийн байгууллагуудыг өөрсдийн салбарын тодорхой шаардлагад нийцүүлэн Загварын хүрээг тохируулахыг дэмжиж, дасан зохицох үйл явцад эдгээр агентлагуудтай хамтран ажиллахыг эрмэлздэг.

**d. Масштаб ба бизнес загвар агностик**

Тодорхой хэмжээ, цар хүрээтэй байгууллагуудад төвлөрөхгүй. Мөн бизнесээс бизнес рүү эсвэл бизнесээс хэрэглэгч рүү чиглэсэн үйл ажиллагаа явуулдаг байгууллагуудад болон бусад ямар ч бизнес загварт ашиглаж болно.

1.2 Хиймэл оюун ухааны ёс зүйтэй ашиглалт, хэрэгжилттэй нягт холбоотой хэд хэдэн асуудал байгааг хүлээн зөвшөөрч байна. Энэхүү Тогтолцооны загварчлал нь эдгээр тусгай асуудлуудыг онцлон авч үздэггүй бөгөөд эдгээр нь тус тусдаа судалгаа, шийдвэр гаргахуйц хэмжээний өргөн хүрээтэй байдаг. Эдгээр асуудлуудын жишээнд дараах зүйлс орно:

- a. Хиймэл оюун ухаанд зориулсан ёс зүйн шинэ зарчмуудыг томъёолох нь энэхүү хувиргах технологийн нарийн төвөгтэй байдлыг даван туулах чухал алхам юм. Дэлхий дахинд бүх нийтийн тогтолцоог бий болгох олон санаачилга хийгдэж байгаа хэдий ч бид соёл, хууль эрх зүйн болон салбарын онцлог хүчин зүйлээс шалтгаалсан ёс зүйн зарчмуудын спектрийг ажиглаж байна. Тогтолцооны загварчлал нь Хавсралт А-д дурдсан ёс зүйн зарчмуудыг ашиглаж, тэдгээрийг хэрэгжүүлэх боломжтой удирдамж болгон хөрвүүлэх замаар энэхүү олон талт байдлыг авч үздэг.
- b. Төрийн болон хувийн хэвшлийн аль нь ч бай байгууллага доторх болон хооронд өгөгдөл хуваалцах асуудлыг шийдвэрлэх нь бат бөх загвар тогтолцоог шаарддаг. IMDA-ийн итгэмжлэгдсэн өгөгдөл хуваалцах тогтолцоо, өгөгдөл хуваалцах мэдээллийн үнэлгээний гарын авлагууд байдаг.
- c. Хиймэл оюун ухаантай холбоотой хууль эрх зүйн хариуцлага, оюуны өмчийн эрх, хиймэл оюун ухааны нийгэмд үзүүлэх нөлөөлөлтэй холбоотой асуудлуудыг хэлэлцэх (жишээ нь, хөдөлмөр эрхлэлт, өрсөлдөөн, хиймэл оюун ухааны бүтээгдэхүүн, үйлчилгээг янз бүрийн сегментүүдийн тэгш бус хүртээмж, хиймэл оюун ухааны технологи буруу хүмүүсийн гарт орох зэрэг) гэх мэт. Эдгээр асуудлууд хэдий ч хамааралтай бөгөөд Сингапурын Менежментийн Их Сургуулийн Хуулийн Сургуульд байгуулагдсан AI болон Өгөгдлийн Засаглалын Төв зэрэг платформуор дамжуулан тусад нь судлах боломжтой.



## 2. ТАНИЛЦУУЛГА

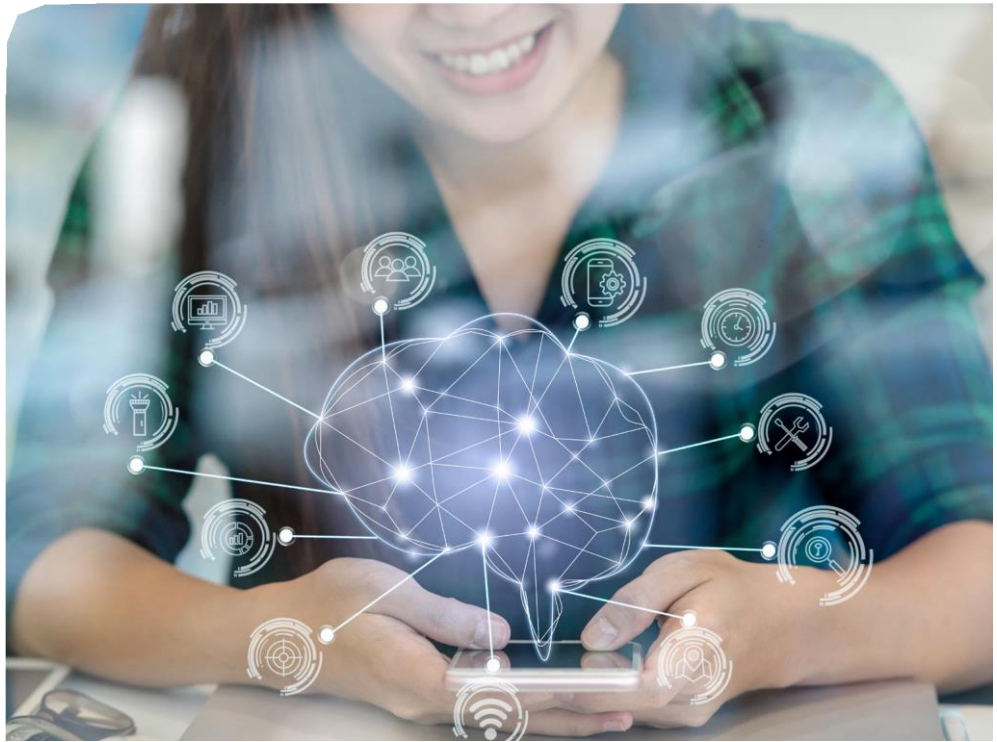


## ЗОРИЛГО

- 2.1 Өгөгдөл болон тооцоолох хүчин чадлын экспоненциал өсөлт нь хиймэл оюун ухаан зэрэг өгөгдөлд суурилсан технологийн дэвшлийг дэмжсэн. Хиймэл оюун ухааныг байгууллагууд шинэ бараа, үйлчилгээ үзүүлэх, бүтээмжийг нэмэгдүүлэх, өрсөлдөх чадварыг нэмэгдүүлэх, улмаар эдийн засгийн өсөлт, амьдралын чанарыг сайжруулахад ашиглах боломжтой. Аливаа шинэ технологийн нэгэн адил хиймэл оюун ухаан нь ёс зүй, хууль эрх зүй, засаглалын шинэ сорилтуудыг бий болгодог. Үүнд шударга бус үр дагаварт хүргэж болзошгүй санамсаргүй ялгаварлан гадуурхах эрсдэл, түүнчлэн тэдний талаар чухал эсвэл эмзэг шийдвэр гаргахад хиймэл оюун ухаан хэрхэн оролцдог талаарх хэрэглэгчдийн мэдлэгтэй холбоотой асуудлууд багтана.
- 2.2 Хувийн Мэдээллийг Хамгаалах Хороо (PDPC)<sup>1</sup> Зөвлөх Зөвлөлийн зөвлөсөний дагуу энэхүү сайн дурын ерөнхий Тогтолцооны загварчлалын хоёр дахь хэвлэлийг AI шийдлүүдийг өргөн хүрээнд нэвтрүүлж буй байгууллагуудад шууд хэрэгжүүлэх боломжтой хэмээн санал болгож байна. Энэхүү Тогтолцооны загварчлал нь хиймэл оюун ухааныг онцлог шинж чанартаа багтаасан арилжааны програм хангамжийн багцуудыг ашиглаж байгаа байгууллагуудад зориулагдаагүй болно.
- 2.3 Энэхүү сайн дурын ерөнхий Тогтолцооны загварчлал нь анхаарах гол асуудал, хэрэгжүүлэх арга хэмжээний талаар зааварчилгааг агуулсан ба үүнийг хэрэгжүүлэхийн тулд байгууллагууд өөрт тохирох эрсдлийг тодорхойлон тохирох арга хэмжээг тодорхойлох шаардлагатай. Тус загварчлал нь байгууллагуудад дараах зорилтуудыг хэрэгжүүлэхэд туслах зорилготой.

- a. Байгууллагууд төрөл бүрийн эрсдэлийг удирдахдаа хиймэл оюун ухааныг ашигласнаар оролцогч талуудын хиймэл оюун ухаанд итгэх итгэлийг бий болгоно.
- b. Дотоод бодлого, бүтэц зохион байгуулалт, өгөгдлийн удирдлага, хамгаалалттай холбоотой процессын хамгийн боломжит үр дүнг харуулах (e.g. the Personal Data Protection Act 2012 (“PDPA”) and the OECD Privacy Principles).

<sup>1</sup> Under section 5 of Singapore’s Personal Data Protection Act 2012, the IMDA is designated as the PDPC.



- 2.4 Тогтолцооны загварчлалыг хэрэгжүүлэх байгууллагуудад туслахын тулд Хувийн мэдээллийг хамгаалах хороо (PDPC) нь нэмэлтээр байгууллагад нэвтрүүлэх, өөрийн үнэлгээ хийх гарын авлагыг (ISAGO-Implementation and self assessment guide for organization)-г бэлтгэсэн. ISAGO нь байгууллагуудад хиймэл оюун ухааны удирдлагын хэрэгжүүлэлт нь Тогтолцооны загварчлалтай нийцэж байгааг үнэлэхэд тусалдаг ба энэ загварчлалд тусгагдсан зөвлөмжийг илүү тодорхой болгох нэмэлт зааварчилгаа, жишээг өгдөг.
- 2.5 Байгууллагууд энэхүү Тогтолцооны загварчлал дахь зөвлөмжийн дагуу хэр хэрэгжүүлэх нь хэд хэдэн хүчин зүйлээс шалтгаална, үүнд байгууллагуудад ашиглах хиймэл оюун ухааны мөн чанар, нарийн төвөгтэй байдал, байгууллагын шийдвэр гаргалтанд хиймэл оюун ухааныг ашиглах, хувь хүмүүсийн бие даасан шийдвэрт нөлөөлөх зэрэг орно.
- 2.6 Тодруулан хэлбэл: AI технологийг хүний шийдвэр гаргалт, эсвэл бие даан шийдвэр гаргахад ашиглаж болно. Жишээлбэл, эмнэлгийн оношлогоонд AI технологийг ашиглан шийдвэр гаргахад үзүүлэх нөлөө нь уламжлалт аргачлалаас илүү сайн байна. Тиймээс хиймэл оюун ухааныг ашиглахад бизнесийн эрсдэл нь хувь хүмүүст үзүүлэх нөлөөлөлтэй пропорциональ байна. Ер нь хиймэл оюун ухааны технологийг хэрэгжүүлэх үеийн зардал нь хүлээгдэж буй үр өгөөжөөсөө давж байвал байгууллагууд хиймэл оюун ухаанаас бусад шийдлүүдийг ашиглах эсэх талаар анхаарч үзэх хэрэгтэй. Энэхүү баримт бичигт тусгагдсан зөвлөмжүүд нь хиймэл оюун ухааны технологийг өргөн хүрээнд үйл ажиллагаандаа нэвтрүүлэхээр шийдсэн байгууллагуудад чиглүүлэх зорилготой юм.





## УДИРДЛАГЫН ЗАРЧИМ

2.7 Тогтолцооны загварчлал нь хиймэл оюун ухаанд итгэх итгэлийг нэмэгдүүлэх, AI технологийн талаарх ойлголтыг сайжруулах зорилготой дээд түвшний хоёр удирдамжийн зарчимд суурилдаг.

- a. Шийдвэр гаргахдаа хиймэл оюун ухааныг ашигладаг байгууллагууд шийдвэр гаргах үйл явц нь ойлгомжтой, **ил тод, шударга** байх ёстой. Төгс тайлбарлах, ил тод байдал, шударга байдлыг хангах боломжгүй ч байгууллагууд хиймэл оюун ухааныг ашиглах, ашиглахдаа эдгээр зарчмын зорилгыг аль болох тусгах байдлаар хэрэгжүүлэхийг хичээх ёстой. Энэ нь хиймэл оюун ухаанд итгэх итгэл, итгэлцлэлийг бий болгоход тусалдаг.
- b. Хиймэл оюун ухааны шийдлүүд нь **хүн төвтэй** байх ёстой.

Хиймэл оюун ухааны шийдлүүд нь хүн төвтэй байх ёстой бөгөөд хүний чадавхийг дээшлүүлэх, хүний ашиг сонирхол, сайн сайхан байдал, аюулгүй байдлыг дизайн, хөгжүүлэлт, ашиглалтын бүх үе шатанд хамгаалахыг чухалчлах ёстой. Хиймэл оюун ухаан эдгээр зарчимд нийцэж байгаа эсэхийг баталгаажуулснаар бид хүний чадавхийг дээшлүүлэх төдийгүй ёс зүйн хэм хэмжээ, нийгмийн үнэт зүйлсийг эрхэмлэх технологийг хөгжүүлэх боломжтой.

*Байгууллагууд хиймэл оюун ухаан ашиглан шийдвэр гаргахдаа ойлгомжтой, ил тод байхын зэрэгцээ шийдлүүд нь хүн төвтэй байх ёстой гэдгийг анхаарах хэрэгтэй.*



- 2.8 Хиймэл оюун ухаан нь бүтээмжийг нэмэгдүүлдэг болон хувь хүн, нийгэм, зах зээл, эдийн засагт нөлөөлж тайлбарлах боломжгүй шийдвэр гаргах боломжтой. Байгууллагууд хиймэл оюун ухааныг нэвтрүүлэх, ёс зүй, ил тод, хариуцлагатай байдал, хэрэглээний шударга байдлыг хангах үүднээс Тогтолцооны загварчлалыг батлах ёстой.
- 2.9 Хиймэл оюун ухааныг өргөн цар хүрээтэй нэвтрүүлэхээр төлөвлөж буй байгууллагууд ёс зүйн тодорхой зарчмуудыг бий болгох ёстой. Тэдгээр нь эмхэтгэсэн хиймэл оюун ухааны ёс зүйн стандартуудыг (Хавсралт А-д байгаа гэх мэт) дурдаж, эдгээр зарчмуудыг одоо байгаа байгууллагын үнэт зүйлстэй нэгтгэж болно. Аюулгүй байдлын асуудал гэх мэт ёс зүйн эрсдэлийг байгууллагын эрсдэлийн хүрээнд удирдах ёстой. Тогтолцооны загварчлал нь ёс зүйн зарчмуудыг засаглалын тогтсон бүтэцтэй уялдуулж, хиймэл оюун ухааныг хариуцлагатай нэвтрүүлэхэд тусалдаг.



## ТААМАГЛАЛ

- 2.10 Тогтолцооны загварчлал нь машин сургалтын загваруудын хүрээнд өгөгдлийн менежментийн сайн туршлагаудыг голчлон авч үздэг. Энэ нь өгөгдлийн менежментийн ерөнхий зарчмуудыг багтаасан модоор удирддаг AI загваруудтай харьцуулахад
- 2.11 Тогтолцооны загварчлал нь хиймэл оюун ухааныг нэвтрүүлэхэд ёс суртахууны болон үйл ажиллагааны суурь үндсийг тогтоодог ч хиймэл оюун ухаантай холбоотой өвөрмөц бөгөөд улам бүр нэмэгдэж буй кибер эрсдлийг шийдвэрлэхэд хангалтгүй юм. Байгууллагууд хиймэл оюун ухааны технологийг ашиглаж байгаа эсэхээс үл хамааран бүтээгдэхүүн, үйлчилгээнийхээ хүртээмж, найдвартай байдал, чанар, аюулгүй байдлыг хамгаалахын тулд кибер аюулгүй байдлын найдвартай арга хэмжээг бие даан хэрэгжүүлэх ёстой.
- 2.12 Энэхүү сайн дурын Тогтолцооны загварчлал баримт бичгийг баталснаар байгууллагыг одоо мөрдөж буй хууль тогтоомжоос чөлөөлөхгүй. Гэсэн хэдий ч энэ нь PDPA, OECD-ийн Нууцлалын зарчим зэрэг стандартуудыг дагаж мөрдөхөд нь тусалдаг, мэдээллийн менежмент, хамгаалалтад хариуцлагад суурилсан практикийг хэрэгжүүлэхэд нь байгууллагуудыг дэмждэг.
- 2.13 Цаашилбал, аж үйлдвэрийн тодорхой салбарууд (санхүү, эрүүл мэнд, хууль эрх зүйн салбар гэх мэт) тухайн салбартай холбоотой одоо мөрдөгдөж буй хууль тогтоомж, дүрэм журам, удирдамжаар зохицуулагдаж болохыг тэмдэглэх нь зүйтэй. Жишээлбэл, Сингапурын санхүүгийн зохицуулах газар хиймэл оюун ухаан, өгөгдөл ашигладаг байгууллагуудад зааварчилгаа өгөх зорилгоор Сингапурын санхүүгийн салбарт хиймэл оюун ухаан, мэдээллийн аналитик ашиглахад шударга байдал, ёс зүй, хариуцлага, ил тод байдлыг дэмжих зарчмуудыг ("FEAT-Fairness, Ethics, Accountability and Transparency зарчмууд") нийтэлсэн. санхүүгийн бүтээгдэхүүн үйлчилгээг санал болгох аналитик Тогтолцооны загварчлал баримт бичгийг батлах нь тухайн байгууллагууд тухайн салбарын хууль тогтоомж, дүрэм журам, удирдамжийг дагаж мөрдөж байна гэсэн үг биш тул эдгээр хууль тогтоомж, дүрэм журам, удирдамжийг анхаарч үзэхийг байгууллагуудад зөвлөж байна.

<sup>2</sup> Monetary Authority of Singapore, "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector" (12 November 2018) <[https://www.mas.gov.sg/publications/monographs\\_or-information-paper/2018/FEAT](https://www.mas.gov.sg/publications/monographs_or-information-paper/2018/FEAT)>.



## ТОДОРХОЙЛОЛТ

2.14 Дараах хялбаршуулсан диаграмм нь Тогтолцооны загварчлалд хиймэл оюун ухааныг нэвтрүүлэх үйл явцын гол оролцогч талуудыг дүрсэлсэн болно. Нэвтрүүлэх үйл явц нь бизнесээс хэрэглэгчид (B2C), бизнесээс бизнесд (B2B), болон бизнесээс бизнесээр дамжуулан хэрэглэгчид (B2B2C) гэсэн харилцааг ялгадаггүй.



Хиймэл Оюун ухааны  
Шийдлийг нийлүүлэгч



Байгууллагууд



Хувь хүн

2.15 Хиймэл оюун ухааны (AI) салбарт хэрэглэгддэг зарим нэр томъёонууд нь тухайн нөхцөл байдал, хэрэглэх хүрээнээс хамааран өөр өөр утгатай байж болно. Энэ Тогтолцооны, тогтолцооны загварчлалд ашиглагддаг зарим гол нэр томъёонуудын тодорхойлолт нь дараах байдалтай байна:

### “AI”

Хиймэл оюун ухаан (AI) нь мэдлэг, оюун ухаан, асуудал шийдвэрлэх, мэдрэхүй, суралцах болон төлөвлөлт зэрэг хүний шинж чанарыг дуурайхыг зорьдог технологийн цогц юм.

Хиймэл оюун ухааны загвараас хамааран, үр дүнд гарсан гаргалгаа эсвэл шийдвэр (урьдчилсан таамаглал, зөвлөмж, ангилал гэх мэт) гаргаж авдаг. AI технологи нь AI алгоритмуудыг ашиглан загваруудыг үүсгэхэд тулгуурладаг. Хамгийн тохиромжтой загваруудыг сонгон авч, үйлдвэрлэлийн системд нэвтрүүлдэг.<sup>2</sup>

### “AI шийдэл нийлүүлэгч”

AI технологийг ашиглан AI шийдлүүд эсвэл хэрэглээний системүүдийг хөгжүүлнэ. Үүнд зөвхөн бэлэн худалдаалагддаг бүтээгдэхүүнүүд, онлайн үйлчилгээ, гар утасны хэрэглээний програмууд болон хэрэглэгчид шууд ашиглаж болох бусад програм хангамж төдийгүй,

B2B2C хэрэглээний програмууд, жишээ нь санхүүгийн байгууллагуудад зарагддаг AI ашигласан луйврын илрүүлэлтийн програм хангамж орно. Үүнд мөн AI ашигласан шинж чанарыг бүтээгдэхүүндээ нэвтрүүлдэг төхөөрөмж, тоног төхөөрөмжийн үйлдвэрлэгчид, бие даасан бүтээгдэхүүн биш боловч эцсийн бүтээгдэхүүнд интеграцчилагдах зориулалттай шийдлүүд орно. Зарим байгууллагууд өөрсдийн AI шийдлүүдийг хөгжүүлж, өөрсдийн шийдлийн ханган нийлүүлэгч болж болно.

### “Байгууллага”

AI технологийг өөрчлөлт, дараах үйл ажиллагаандаа ашиглаж байгаа байгууллагуудыг илэрхийлдэг.

Үйл ажиллагаа (зээл авах аппликейшны процесс гм) байгууллагын харилцагч нартай хамтран ажиллах үйлчилгээнүүд (онлайн худалдааны портал, AI функц, модулийг ашиглан борлуулалт түгээлт (гэрийн ухаалаг удирдлагын систем).

### “Хувь хүн”

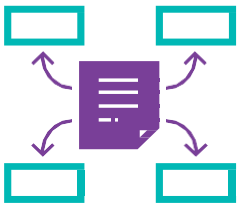
Байгууллагад хиймэл оюун ухааны бүтээгдэхүүн ба/эсвэл үйлчилгээ нийлүүлэх гэж буй хүмүүс эсвэл хиймэл оюун ухааны бүтээгдэхүүн үйлчилгээг

аль хэдийн худалдаж авсан хүмүүсийг нэрлэж болно. Эдгээрийг мөн "хувь хүн" эсвэл "хэрэглэгч" гэж нэрлэж болно.

<sup>2</sup> AI-ийн энэхүү тодорхойлолтыг янз бүрийн эх сурвалжаас авч, энэ загварын хүрээнд тохируулан ашигласан болно. Үүнийг албан ёсны эсвэл бүрэн гүйцэд тодорхойлолт гэж үзэж болохгүй.



### **3. ХИЙМЭЛ ОЮУН УХААНЫ ЗАСАГЛАЛЫН ТОГТОЛЦООНЫ ЗАГВАРЧЛАЛ**



## ХИЙМЭЛ ОЮУН УХААНЫ ЗАСАГЛАЛЫН ТОГТОЛЦООНЫ ЗАГВАРЧЛАЛ

3.1 Энэхүү тогтолцооны загварчлал нь байгууллагуудын AI -ийн хариуцлагатай хэрэглээг дэмжих арга хэмжээнүүдийг дараах гол чиглэлүүдэд хэрэгжүүлэхэд зориулсан зөвлөмжийг агуулна.



**a. Дотоод удирдлагын бүтэц болон арга хэмжээнүүд**

Одоогийн дотоод удирдлагын бүтэц болон арга хэмжээнүүдийг өөрчлөх эсвэл шинээр бий болгохдоо алгоритмын шийдвэр гаргалттай холбоотой үнэт зүйлс, эрсдэлүүд, хариуцлагыг тусгах.



**b. Хиймэл оюун ухаан ашигласан шийдвэр гаргалтанд хүний оролцооны түвшинг тодорхойлох**

Байгууллагуудад хиймэл оюун ухаан ашиглах эрсдэлийн хязгаарыг тогтоох арга зүй, өөрөөр хэлбэл, хүлээн авах боломжтой эрсдэлийг тодорхойлох, хиймэл оюун ухаан ашигласан шийдвэр гаргалтад хүний оролцооны тохиромжтой түвшнийг тогтоох,



**c. Үйл ажиллагааны удирдлага**

Хиймэл оюун ухааны загваруудыг боловсруулах, сонгох болон хадгалах үед анхаарах асуудлууд, өгөгдлийн удирдлагыг мөн хамааруулна.



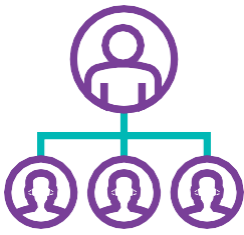
**d. Оролцогч талуудын хамтын ажиллагаа**

Байгууллагын оролцогч талуудтай харилцах стратеги, тэдэнтэй харилцах харилцааг удирдах арга замуудыг хамааруулна.

3.2 Байгууллагууд энэхүү загварын фрэймворкыг хэрэгжүүлэхдээ зарим элементүүд нь тухайн байгууллагад ач холбогдолгүй болохыг харна. Энэ загварын фрэймворк нь уян хатан байдлаар хийгдсэн бөгөөд байгууллагууд өөрсдийн хэрэгцээнд нийцүүлэн өөрчилж, холбогдох элементүүдийг сонгон авч болох юм.

3.3 Байгууллагуудад загварын фрэймворкыг илүү сайн ойлгоход нь туслахын тулд тус бүр хэсэгт байгууллагууд практик дээр хэрхэн хэрэгжүүлсэн болохыг харуулсан жишээнүүдийг оруулсан. Мөн PDPC нь орон нутгийн болон олон улсын төрөл бүрийн байгууллагууд загварын фрэймворкын бүх хэсгүүдэд нийцсэн хиймэл оюун ухааны удирдлагын практикыг хэрхэн хэрэгжүүлсэн болохыг харуулсан Хэрэглээний жишээнүүдийн товхимлыг гаргасан.





## ДОТООД ЗАСАГЛАЛЫН БҮТЭЦ БОЛОН АРГА ХЭМЖЭЭНҮҮД

- 3.4 Энэ хэсэг нь байгууллагуудад хиймэл оюун ухааны технологийг үйл ажиллагаа болон/эсвэл бүтээгдэхүүн, үйлчилгээндээ хэрхэн нэвтрүүлэх талаар зохих хяналтыг хэрэгжүүлэх боломжийг олгодог дотоод засаглалын бүтэц зохион байгуулалтыг хөгжүүлэхэд чиглүүлэх зорилготой.
- 3.5 Дотоод засаглалын бүтэц, арга хэмжээнүүд нь байгууллагын хиймэл оюун ухааны хэрэглээнд хүчтэй хяналт тавихад тусалдаг. Байгууллагын одоо байгаа дотоод засаглалын бүтцийг өөрчлөн тохируулах эсвэл шаардлагатай тохиолдолд шинэ бүтцийг нэвтрүүлэх боломжтой. Жишээлбэл, хиймэл оюун ухааны хэрэглээтэй холбоотой эрсдлийг байгууллагын эрсдлийн удирдлагын бүтэц дотор зохицуулж болох бол ёс зүйн асуудлуудыг байгууллагын үнэт зүйлс болгон нэвтрүүлж, ёс зүйн хяналтын зөвлөл эсвэл ижил төстэй бүтэцээр дамжуулан удирдаж болно.

*Ёс зүйн асуудлуудыг байгууллагын үнэт зүйлс болгон нэвтрүүлж, ёс зүйн хяналтын зөвлөл эсвэл ижил төстэй бүтэцээр удирдах боломжтой.*

- 3.6 Байгууллагууд дотоод засаглалын бүтцэд тохирох онцлог шинжүүдийг тодорхойлох талаар авч үзэх нь зүйтэй. Жишээлбэл, төвлөрсөн засаглалын механизмд бүрэн найдах нь тохиромжгүй бол, ёс зүйн асуудлуудыг өдөр тутмын шийдвэр гаргалтад тусгах зорилгоор үйл ажиллагааны түвшинд төвлөрсөн бус механизмийг авч үзэж болно. Байгууллагын хиймэл оюун ухааны засаглалд байгууллагын дээд удирдлага болон удирдах зөвлөлийн ивээн тэтгэлэг, дэмжлэг, оролцоо маш чухал юм.
- 3.7 Байгууллагууд дотоод засаглалын бүтцийн хөгжүүлэлтэд хамааралтай дараах онцлог шинжүүдийг багтаах талаар авч үзэж болно:

## 1. Хиймэл оюун ухааныг ёс зүйтэйгээр нэвтрүүлэхэд тодорхой үүрэг, хариуцлагыг тогтоох:

- a. Хиймэл оюун ухааныг нэвтрүүлэхтэй холбоотой төрөл бүрийн үе шат, үйл ажиллагааны хариуцлага, хяналтыг зохих албан тушаалтан болон/эсвэл хэлтэсүүдэд хуваарилах ёстой. Шаардлагатай болон боломжтой тохиолдолд холбогдох мэргэжилтнүүдээс бүрдсэн, байгууллагын бүх хэсгээс зохих төлөөлөлтэй зохицуулах байгууллагыг байгуулах талаар авч үзэх хэрэгтэй.
- b. Дотоодын хиймэл оюун ухааны засаглалын үүрэгтэй ажилтнууд болон/эсвэл хэлтэсүүд өөрсдийн үүрэг, хариуцлагыг бүрэн мэдэж, зохих ёсоор сургалтанд хамрагдаж, үүргээ гүйцэтгэхэд шаардлагатай нөөц, удирдамжаар хангагдсан байх ёстой.
- c. Тодорхойлж болох гол үүрэг, хариуцлагууд:
  - i. Эрсдэлийн удирдлагын аливаа тогтолцоог ашиглах, эрсдэлийн хяналтын арга хэмжээг хэрэгжүүлэх ("Эрсдэлийн удирдлага ба дотоод хяналт"-ыг доороос үзнэ үү):
    - o Хиймэл оюун ухааныг ашиглах эрсдэл, түүний дотор хувь хүмүүст үзүүлэх сөрөг нөлөөллийг үнэлж, удирдах (**Хамгийн эмзэг бүлгүүд хэн бэ**, Тэдэнд хэрхэн нөлөөлж байна, Нөлөөллийн цар хүрээг хэрхэн үнэлэх, Нөлөөлөлд өртсөн бүлгүүдээс хэрхэн санал авах гм).
    - o Хүний оролцооны тохиромжтой түвшинг тодорхойлох
    - o Хиймэл оюун ухааны загварын сургалт болон сонголтын процессыг удирдах

- ii. Нэвтрүүлсэн хиймэл оюун ухааны загваруудын засвар үйлчилгээ, хяналт, баримтжуулалт, үнэлгээ хийх:
- iii. Харилцаа холбооны суваг болон оролцогч талуудтай хийж буй харилцааг үнэлэх, ил тод байдал болон үр дүнтэй санал хүсэлтийн сувгуудыг хангах:
- iv. Хиймэл оюун ухааны системүүдтэй харьцаж буй холбогдох ажилтнуудыг зохих ёсоор сургах Шаардлагатай тохиолдолд, хиймэл оюун ухааны загваруудтай шууд ажилладаг, харьцдаг ажилтнуудыг дараах чиглэлүүдээр сургах шаардлагатай.  
Эдгээр сургалтын чиглэлүүдийг хэрэгжүүлснээр ажилтнууд хиймэл оюун ухааны загваруудыг үр дүнтэй ашиглаж, гарсан шийдвэрийг ойлгож, тайлбарлаж, өгөгдлийн алдааг удирдаж, ёс зүйтэй ажиллах чадварыг дээшлүүлнэ.

## 2. Эрсдлийн удирдлага ба дотоод хяналт

- a. Байгууллагууд сонгосон хиймэл оюун ухааны загварыг нэвтрүүлэхтэй холбоотой эрсдлүүдийг шийдвэрлэхэд чиглэсэн эрсдлийн удирдлага болон дотоод хяналтын найдвартай системийг хэрэгжүүлэх талаар авч үзэж болно.
- b. Ийм арга хэмжээнүүд нь дараах зүйлсийг багтааж болно:
  - i. Хиймэл оюун ухааны загварын сургалтад ашигласан өгөгдлийн багцууд зорилгодоо нийцсэн, хангалттай байхын тулд зохих хүчин чармайлт гаргаж, өгөгдлийн нарийвчлалгүй байдал эсвэл туйлшралын эрсдлийг үнэлэх, удирдах, мөн загварын сургалтын явцад илэрсэн алдаануудыг хянах. Бараг бүх өгөгдлийн багц ямар нэг хэмжээгээр туйлширсан байдаг. Байгууллагууд өгөгдлийн багцуудын туйлшралын шалтгааныг ойлгож, үүнийг аюулгүй байдлын арга хэмжээ болон нэвтрүүлэлтийн стратегидээ тусгахыг хичээх хэрэгтэй
  - ii. Нэвтрүүлсэн хиймэл оюун ухааны гүйцэтгэл болон бусад асуудлуудыг зохих түвшний удирдлагад мэдэгдэх хяналт, тайлагналтын систем болон үйл явцыг бий болгох. Зохих тохиолдолд, хяналт нь хүний хяналтыг үр дүнтэй өргөжүүлэхийн тулд автомат хяналтыг багтааж болно. Хиймэл оюун ухааны системийг өөрийн таамаглалын итгэлийн түвшинг тайлагнах, мөн тухайн түвшинд яагаад итгэж байгааг тайлбарлах боломжтой байдлаар зохион бүтээж болно.
  - iii. Хиймэл оюун ухаантай холбоотой үйл ажиллагаанд оролцдог гол ажилтнууд солигдох үед мэдлэгийг хуваалцах асуудлыг зохих ёсоор хангах. Энэ нь ажилтнуудын хөдөлгөөнөөс үүдэн дотоод засаглалд хоосон зай үүсэх эрсдэлийг бууруулна.
  - iv. Байгууллагын бүтэц эсвэл хиймэл оюун ухаантай холбоотой гол ажилтнуудад томоохон өөрчлөлт гарсан үед дотоод засаглалын бүтэц болон арга хэмжээг хянах.
  - v. Дотоод засаглалын бүтэц болон арга хэмжээнүүдийг тогтмол хянаж, тэдгээрийн үр ашиг, зохицлыг үргэлжлүүлэн хангах.

# CUJO AI:

## ДОТООД ЗАСАГЛАЛЫН БҮТЭЦ БОЛОН АРГА ХЭМЖЭЭНИЙ ЗУРАГЛАЛ

CUJO AI нь харилцаа холбооны операторуудын зах зээлд үйл ажиллагаа явуулдаг сүлжээний ухаалаг програм хангамжийн компани юм. АНУ-д төвтэй бөгөөд өрхийн болон бизнесийн холболтын төхөөрөмжүүдийн аюулгүй байдал, хяналт, нууцлалыг сайжруулахын тулд хиймэл оюун ухааныг хөгжүүлж, нэвтрүүлэхийг зорьж байна.

CUJO AI нь хиймэл оюун ухааныг ашиглахдаа хатуу хяналт тавихын тулд тодорхой дотоод удирдлагын бүтэц, арга хэмжээг хэрэгжүүлсэн. Үүний олон талын оролцоотой удирдлагын бүтэц нь зохих түвшинд шийдвэр гаргахад тусалдаг:

### A Research Board

Ерөнхий Технологи Эрхэлсэн Захирал, Лабораторийн Дарга болон Ерөнхий Өгөгдөл Судлаачийн бүрэлдэхүүнтэй баг хиймэл оюун ухааны хөгжүүлэлт болон нэвтрүүлэлтийг баталдаг.

Тухайлбал, Ерөнхий Технологи Эрхэлсэн Захирал нь 100 гаруй ажилтантай дөрвөн техникийн багийг удирддаг.

**Тэдний үүрэг, хариуцлагуудыг тодорхой заасан байдаг:**

- Судалгааны баг нь Өгөгдлийн шинжилгээ хийх, Судалгаа явуулах, Хиймэл оюун ухааны алгоритмууд болон Машин сургалтын загваруудыг хөгжүүлдэг ("ML");
- Инженерийн баг нь Програм хангамж бүтээх, Үүлэн үйлчилгээ хөгжүүлэх, Аппликейшнүүд боловсруулдаг
- Үйл ажиллагааны баг нь Хиймэл оюун ухааны загварыг нэвтрүүлэх, Платформыг шинэчлэх ажлыг хийдэг;
- Хүлээлгэн өгөх баг нь Операторуудтай хамтран ажиллах, Үйлчилгээг интеграцчилах ажлыг зохион байгуулдаг.

### An Architecture Steering Group ("ASG")

Ерөнхий Технологи Эрхэлсэн Захирал, Ерөнхий Архитектурын Захирал болон ахлах инженерүүдийн бүрэлдэхүүнтэй баг хиймэл оюун ухаан, машин сургалтын загваруудын бат бөх байдлыг нэвтрүүлэхээс өмнө хангадаг. Энэхүү ASG баг нь хоёр долоо хоног тутамд хуралдаж, судалгааны баг машин сургалтын загварууд болон хиймэл оюун ухааны алгоритмуудын талаархи судалгааны үр дүнгээ хуваалцдаг.

### PhD-level employees

Хиймэл оюун ухааны хөгжүүлэлт, нэвтрүүлэлтийн процессыг хянаж, шинэ функц хөгжүүлэх бүрт академик хяналтын стандартыг хэрэгжүүлэхийг зорьдог.

Үүнээс гадна, CUJO AI нь ажилтнууддаа зориулан "Ёс зүйн код" ("Код") боловсруулсан. Шинэ ажилтнуудыг CUJO AI-ийн орон нутгийн баримт бичиг болон процессын сантай танилцуулдаг. Жишээлбэл, CUJO AI-ийн Финланд дахь оффис ажилтнууддаа электрон "CUJO ажилтны гарын авлага"-аар хангадаг. Гарын авлага нь Ёс зүйн кодыг дэлгэрэнгүй тайлбарлахын зэрэгцээ бизнесийн ёс зүй, зан төлөв зэрэг бусад сэдвүүдийг хамардаг. Ажилтнууд дараах **ёс зүйн зарчмуудыг** үндэслэн өөрсдийн үүрэг, хариуцлагыг хэрэгжүүлдэг:

- a. Дэлхийн өнцөг булан бүрт байрлах салбар оффисуудаараа дамжуулан бизнесийг шударга, ёс зүйтэйгээр явуулах;
- b. Шийдвэр гаргахдаа үнэнч байдал, шударга байдал, хүндэтгэл, хариуцлага, ёс зүй, итгэлцэл, болон оновчтой бизнесийн шүүлтэд үндэслэх.
- c. Оффисын ажилтанууд, захирлууд, болон түншүүдийн хууль бус эсвэл ёс зүйгүй үйлдэл нь компанийн ашиг сонирхолд нийцэхгүй болохыг тодорхойлох;
- d. Богино хугацааны ашиг сонирхлын төлөө компанийн зарчмуудыг золиослохгүй байх.

# MASTERCARD:

## ДОТООД ЗАСАГЛАЛЫН БҮТЭЦ БОЛОН АРГА ХЭМЖЭЭНИЙ ЗУРАГЛАЛ

Mastercard нь дэлхийн төлбөрийн системийн салбарт үйл ажиллагаа явуулдаг технологийн компани юм. Түүний дэлхийн төлбөрийн системийн боловсруулалтын сүлжээ нь 210 гаруй улс орон, нутаг дэвсгэрт хэрэглэгчид, санхүүгийн байгууллагууд, худалдаачид, засгийн газрууд болон бизнесүүдийг холбодог. Mastercard нь өөрийн үзэл баримтлалаа хэрэгжүүлэхийн тулд хиймэл оюун ухааныг залилангаас урьдчилан сэргийлэх, ирээдүйн санхүүгийн чиг хандлагыг таамаглах, хэрэглэгчийн жижиглэн худалдааны туршлагыг сайжруулах зэрэг олон талбарт ашигладаг.

Mastercard нь хиймэл оюун ухааныг ашиглахдаа нарийн чанд хяналтыг хангахдаа өндөр эрсдэлтэй гэж тодорхойлсон AI програмуудыг хэрэгжүүлэлтийг хянаж, батлахын тулд Засаглалын Зөвлөл (Governance Council) байгуулсан. Засаглалын Зөвлөлийг Хиймэл Оюун Ухааны Төвийн Гүйцэтгэх Дэд Ерөнхийлөгч удирддаг бөгөөд зөвлөлийн гишүүд нь Ерөнхий Өгөгдөл Хариуцсан Ажилтан, Ерөнхий Нууцлал Хариуцсан Ажилтан, Ерөнхий Мэдээллийн Аюулгүй Байдлын Ажилтан, өгөгдлийн судлаачид болон бизнес багуудын төлөөллөөс бүрддэг.

Mastercard нь Засаглалын Зөвлөлд **тодорхой үүрэг, хариуцлагыг тодорхойлсон** ба Зөвлөлийн төлөөлөгч бүр өөрийн мэргэжлийн ур чадварыг шийдвэр гаргах үйл явцад авчирдаг:

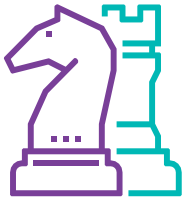


- a. **Chief Data Officer and Chief Privacy Officer**
- Зөвлөл хиймэл оюун ухааныг хэрэгжүүлэх саналыг хянаж, дараах зүйлсийг хангана:
- Өгөгдөл нь хиймэл оюун ухааны зорилгод тохиромжтой байх;
  - Хиймэл оюун ухааныг ёс зүйтэй зорилгоор ашигласан байх
  - Хувь хүнд үзүүлэх нөлөө нь тохиромжтой байх бөгөөд боломжит хохирол (нууцлал болон өгөгдөл хамгаалалтад учирч болох эрсдэлийг оролцуулан) хангалттай бууруулагдсан байх.
- b. **Chief Information Security Officer**
- Аюулгүй байдлыг загварчлалаас нь эхлэн хэрэгжүүлж байгааг хангах.
- c. **Data Science teams**
- Хиймэл оюун ухаан бүтээж, хэрэгжүүлж буй багууд Өгөгдлийн Алба болон Нууцлалын Албатай тогтмол харилцаатай байх ба хиймэл оюун ухааны програм хэрэгжүүлэлтийн i-шаардлагатай засаглалын болон самьдралын мөчлөгийн талаар мэдээллийг тогтмол хуваалцаж байх.

Mastercard нь хиймэл оюун ухааны нэвтрүүлэлттэй холбоотой эрсдлийг шийдвэрлэхийн тулд эрсдлийн удирдлага болон дотоод хяналтыг хэрэгжүүлсэн.

Жишээлбэл, Mastercard нь санал болгож буй хиймэл оюун ухааны үйл ажиллагааны эрсдлийг тодорхойлохын тулд анхан шатны эрсдлийн үнэлгээг явуулдаг бөгөөд үүнд олон хүчин зүйлийг үнэлдэг.

Үүнээс гадна, Mastercard нь цуглуулж буй өгөгдлөөс үүдэлтэй эрсдэл эсвэл үйл ажиллагааны боломжит туйлшралын түвшинг бууруулах боломжит арга хэмжээг тодорхойлно. Хэрэв хиймэл оюун ухааны төсөл өндөр эрсдэлтэй гэж тодорхойлогдвол Засаглалын Зөвлөлд хянуулахаар шилжүүлнэ. Харин бага эрсдэлтэй төслүүд хяналтад хамрагдахгүй бөгөөд шууд загвар хөгжүүлэлтийн шатанд шилжиж болно.



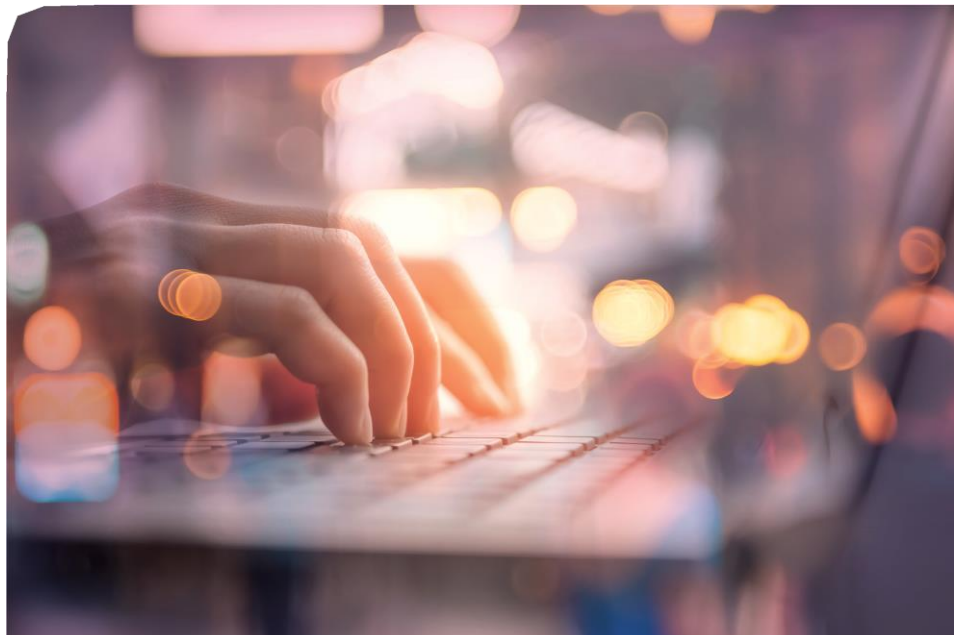
## ХИЙМЭЛ ОЮУН УХААНЫ ТУСЛАМЖТАЙ ШИЙДВЭР ГАРГАЛТАД ХҮНИЙ ОРОЛЦООНЫ ТҮВШНИЙГ ТОГТООХ НЬ

- 3.8 Энэ хэсэг нь байгууллагуудад хиймэл оюун ухааны тусламжтай шийдвэр гаргалтад хүний хяналтыг зохистой хэмжээгээр тодорхойлоход туслах зорилготой юм.
- 3.9 Хиймэл оюун ухаан ашиглах зорилгыг тодорхой болгох нь хүний хяналтыг тодорхойлоход анхны чухал алхам юм. Байгууллагууд хиймэл оюун ухаан ашиглах арилжааны зорилгоо шийдэхээс эхэлж болно (жишээ нь, шийдвэр гаргалтын тууштай байдлыг хангах, үйл ажиллагааны үр ашгийг дээшлүүлэх, зардлыг бууруулах, эсвэл хэрэглэгчдийн сонголтыг нэмэгдүүлэх шинэ бүтээгдэхүүний онцлог шинжийг нэвтрүүлэх). Эдгээр арилжааны зорилгуудыг байгууллагын шийдвэр гаргалтад хиймэл оюун ухаан ашиглах эрсдэлтэй харьцуулан үнэлж болно. Энэхүү үнэлгээг байгууллагуудын компанийн үнэт зүйлсээр удирдан чиглүүлж, улмаар байгууллагууд үйл ажиллагаа явуулж буй газар нутгийн нийгмийн хэм хэмжээ эсвэл хүлээлтийг тусгах боломжтой.

*Хиймэл оюун ухааны шийдлүүдийг нэвтрүүлэхээс өмнө байгууллагууд хиймэл оюун ухааныг ашиглах арилжааны зорилгоо тодорхойлж, дараа нь эдгээрийг байгууллагын шийдвэр гаргалтад хиймэл оюун ухаан ашиглахтай холбоотой эрсдэлтэй харьцуулан үнэлэх ёстой.*

- 3.10 Олон улс оронд үйл ажиллагаа явуулдаг байгууллагуудын хувьд, аль болох нийгмийн хэм хэмжээ, үнэт зүйлс болон/эсвэл хүлээлтийн ялгааг харгалзан үзэх нь зүйтэй. Жишээлбэл, тоглоомын сурталчилгаа нэг улсад хүлээн зөвшөөрөгдөж байхад өөр улсад хүлээн зөвшөөрөгдөхгүй байж болно. Нэг улсын дотор ч гэсэн хиймэл оюун ухаан ашиглаж буй орчноос хамааран эрсдэл ихээхэн ялгаатай байж болно. Жишээлбэл, онлайн худалдааны төвд бүтээгдэхүүн санал болгодог зөвлөмжийн систем эсвэл аялалын даатгалын онлайн өргөдлийг автоматаар батлах нь мэргэжлийн хөрөнгө оруулагчдад санал болгодог алгоритмын арилжааны үйлчилгээтэй холбоотой эрсдэлээс бага байж болно.

- 3.11 Зарим эрсдэлүүд хувь хүмүүст зөвхөн бүлгийн түвшинд илэрч болно. Жишээлбэл, хувьцааны зөвлөмжийн алгоритмыг өргөн хүрээтэй ашиглах нь олон хүн нэгэн зэрэг ижил шийдвэр гаргаснаар зах зээлийн бүхэлд нь хэлбэлзэл нэмэгдэх сүргийн зан үйлийг үүсгэж болно. Хувь хүмүүст учирч болох эрсдэлээс гадна бусад төрлийн эрсдэлүүдийг ч мөн тодорхойлох боломжтой (жишээ нь, байгууллагын арилжааны нэр хүндэд учирч болох эрсдэл).
- 3.12 Байгууллагуудын арилжааны зорилгоо хиймэл оюун ухаан ашиглах эрсдэлтэй харьцуулах үйл явцыг аль болох компанийн үнэт зүйлсээр удирдан чиглүүлэх нь зүйтэй. Байгууллагууд хиймэл оюун ухааныг нэвтрүүлэх болон алгоритмын шийдвэр гаргалтанд ашиглах загвар нь өөрсдийн үндсэн үнэт зүйлстэй нийцэж байгаа эсэхийг үнэлэх боломжтой. Ямар нэгэн зөрчил, гажуудал гарсан тохиолдолд байгууллагууд тодорхой, баримтжуулсан үндэслэлтэйгээр ухамсартай шийдвэр гаргах ёстой.
- 3.13 Арилжааны зорилго, эрсдэлүүдийг тодорхойлох болон хиймэл оюун ухааны тусламжтай шийдвэр гаргалтад хүний оролцооны зохистой түвшинг тодорхойлох нь давтагдах болон тасралтгүй үргэлжлэх үйл явц тул байгууллагууд технологийн шийдэлтэй холбоотой эрсдэлүүдийг тогтмол тодорхойлж, үнэлж, тэдгээрийг бууруулан, хэрэв бууруулах арга хэмжээ амжилтгүй болсон тохиолдолд хариу арга хэмжээ авах төлөвлөгөөтэй байх нь зүйтэй. Энэ үйл явцыг үе үе хянаж үздэг эрсдэл нөлөөллийн үнэлгээгээр баримтжуулснаар байгууллагууд хиймэл оюун ухааны шийдлүүдийг ашиглахад илүү тодорхой, итгэлтэй болно. Мөн энэ нь байгууллагуудад хувь хүмүүс, бусад байгууллага, бизнес эсвэл зохицуулагчдаас үүсэх боломжит сорилтуудад хариу өгөхөд тусална.



## ХИЙМЭЛ ОЮУН УХААН АШИГЛАН ШИЙДВЭР ГАРГАЛТАД ХҮНИЙ ОРОЛЦООГ ГУРВАН ӨРГӨН ХҮРЭЭТЭЙ БАЙДЛААР АНГИЛЖ БОЛНО:

3.14 Дээр дурдсан эрсдлийн удирдлагын хандлагад үндэслэн, Загвар Хүрээ нь шийдвэр гаргах үйл явц дахь хүний хяналтын янз бүрийн түвшинг ангилахын тулд гурван өргөн хүрээний арга барилыг тодорхойлсон байна:

- a. **Хүний бүрэн хяналттай шийдвэр гаргалт:** Энэ тохиолдолд хиймэл оюун ухааны систем нь зөвхөн мэдээлэл болон зөвлөмж өгөх үүрэгтэй бөгөөд эцсийн шийдвэрийг хүн гаргана. Энэ нь өндөр эрсдэлтэй шийдвэр гаргах үйл явцад тохиромжтой бөгөөд хүний хяналт хамгийн өндөр түвшинд байдаг.

Жишээлбэл, эмч хиймэл оюун ухааныг ашиглан танил бус эмнэлгийн нөхцлийн боломжит оношилгоо болон эмчилгээг тодорхойлж болно. Гэсэн хэдий ч, эцсийн онош болон түүнд тохирох эмчилгээг эмч гаргана. Энэ загвар нь хиймэл оюун ухааныг хүний мэдээлэлд суурилсан шийдвэр гаргахад хангалттай мэдээлэл өгөхийг шаарддаг (жишээ нь, шийдвэрт ашигласан хүчин зүйлс, тэдгээрийн утга болон жингүүд, хамаарал).

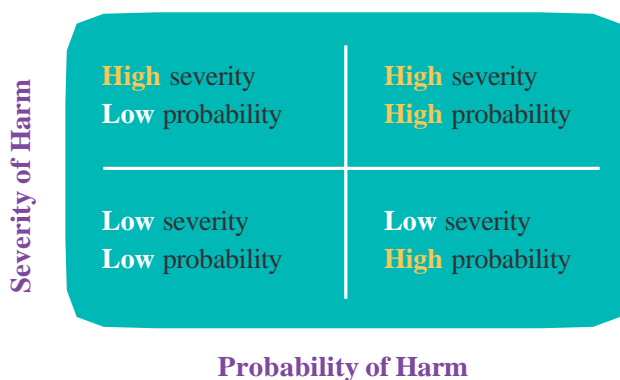
- b. **Хүний оролцоогүй (Human-out-of-the-loop):** Энэ нь шийдвэр гүйцэтгэлийн явцад хүний хяналт байхгүй гэдгийг илэрхийлнэ. Хиймэл оюун ухааны систем нь хүний оролцоогүйгээр бүрэн хяналтыг авч, шийдвэр гаргах эрхтэй.

Жишээлбэл, бүтээгдэхүүний зөвлөмжийн шийдэл нь урьдчилан тогтоосон демографик болон зан төлөвийн профайлуудад үндэслэн хувь хүмүүст бүтээгдэхүүн болон үйлчилгээг автоматаар санал болгож болно. Хиймэл оюун ухаан нь мөн динамик байдлаар шинэ профайлуудыг бүтээж, дараа нь бүтээгдэхүүн болон үйлчилгээг санал болгоход ашиглаж болно.

Машин сургалтын загварыг мөн нислэгийн компаниуд ирээдүйн эрэлт эсвэл боломжит тасалдлуудыг урьдчилан таамаглахад ашиглаж болно. Энэ загварын үр дүнг нислэгийн хуваарийг оновчтой болгоход ашигладаг шийдлийн модульд ашигладаг бөгөөд энэ нь хүний оролцоогүйгээр хийгддэг.

- c. **Хүний хяналттай (Human-over-the-loop эсвэл human-on-the-loop):** Энэ нь хүний хяналт нь хянагч эсвэл хяналтын үүрэгтэй байдгийг илэрхийлнэ. Хүн нь хиймэл оюун ухааны загвар нь хүлээгдээгүй эсвэл хүсээгүй үйл явдалтай тулгарсан үед хяналтыг авах боломжтой байдаг (жишээлбэл, загварын алдаа гарсан үед). Энэ арга нь алгоритмын үйл ажиллагааны явцад хүнийг параметруудийг тохируулах боломжийг олгодог. Жишээлбэл, GPS навигацийн систем нь А цэгээс Б цэг хүртэл замыг төлөвлөж, жолоочид сонгоход зориулсан хэд хэдэн боломжит маршрутыг санал болгодог. Жолооч аяллын явцад (жишээлбэл, гэнэтийн замын түгжрэлээс болж) параметруудийг өөрчилж болно, ингэснээр маршрутыг дахин программчлах шаардлагагүй.

- 3.15 Загварын хүрээ нь мөн хиймэл оюун ухааныг сайжруулсан шийдвэр гаргахад шаардагдах хүний оролцооны түвшинг тодорхойлоход туслах дизайны тогтолцоог (матриц хэлбэрээр зохион байгуулсан) санал болгодог. Энэхүү дизайны хүрээ нь хоёр тэнхлэгийн дагуу бүтэцлэгдсэн: (а) магадлал; (б) тухайн хувь хүн (эсвэл байгууллагын) талаар байгууллагаас гаргасан шийдвэрийн үр дүнд хувь хүн (эсвэл байгууллагад) учирсан хохирлын ноцтой байдал.
- 3.16 "Хохирол" гэсэн тодорхойлолт, мөн магадлал болон ноцтой байдлыг тооцоолох нь нөхцөл байдлаас хамаарч, салбар бүрт өөр өөр байна. Жишээлбэл, эмнэлгийн хувьд өвчтөний эмнэлгийн байдлын буруу оношилгоотой холбоотой хохирлыг авч үзэх асуудлууд нь хувцасны дэлгүүрийн хувцасны буруу бүтээгдэхүүний зөвлөмжөөс үүдэлтэй хохирлыг авч үзэх асуудлуудаас ялгаатай байна.



- 3.17 Гэсэн хэдий ч, матрицыг байгууллагын хиймэл оюун ухаан ашигласан шийдвэр гаргалтын үйл явц дахь хүний хяналтын түвшинг тодорхойлоход зөвхөн хохирлын магадлал болон хохирлын ноцтой байдлыг авч үзэх цорын ганц хүчин зүйл гэж үзэж болохгүй (хэдийгээр эдгээр нь ерөнхийдөө хамгийн чухал хоёр хүчин зүйл юм).<sup>4</sup>
- 3.18 Аюулгүй байдалд чухал нөлөөтэй системүүдийн хувьд байгууллагууд хүний хяналт авах боломжийг хангах нь зүйтэй бөгөөд хиймэл оюун ухааны систем нь тухайн хүнд утга учиртай шийдвэр гаргах эсвэл хүний хяналт боломжгүй үед системийг аюулгүй зогсоох хангалттай мэдээлэл өгөх ёстой.

<sup>4</sup> Байгууллагуудын янз бүрийн нөхцөлд авч үзэж болох бусад хүчин зүйлүүд нь:

- Хохирлын шинж чанар (жишээ нь, хохирол нь биет эсвэл биет бус эсэх);
- Хохирлын эргэлт (буцаах боломж), мөн үүнтэй холбоотойгоор хувь хүмүүст нөхөн олговор авах боломж;
- Хүнийг шийдвэр гаргах үйл явцад оролцуулах нь үйл ажиллагааны хувьд боломжтой эсвэл утга учиртай эсэх (жишээ нь, өндөр хурдтай санхүүгийн арилжааны үед хүний оролцоог хангах нь боломжгүй бөгөөд жолоочгүй тээврийн хэрэгслийн хувьд үр дүнгүй байх).

## ХОХИРЛЫН МАГАДЛАЛ-НОЦТОЙ БАЙДЛЫН МАТРИЦИЙГ АШИГЛАХ НЬ

Онлайн жижиглэнгийн дэлгүүр нь хэрэглэгчдийн үзэж байсан бүтээгдэхүүнүүд болон худалдан авалтын түүхэнд үндэслэн хүнсний бүтээгдэхүүнийг бүрэн автоматаар санал болгоход хиймэл оюун ухааныг ашиглахыг хүсэж байна. Энэхүү автоматжуулалт нь байгууллагын үйл ажиллагааны үр ашгийг дээшлүүлэх арилжааны зорилготой нийцэж байна.

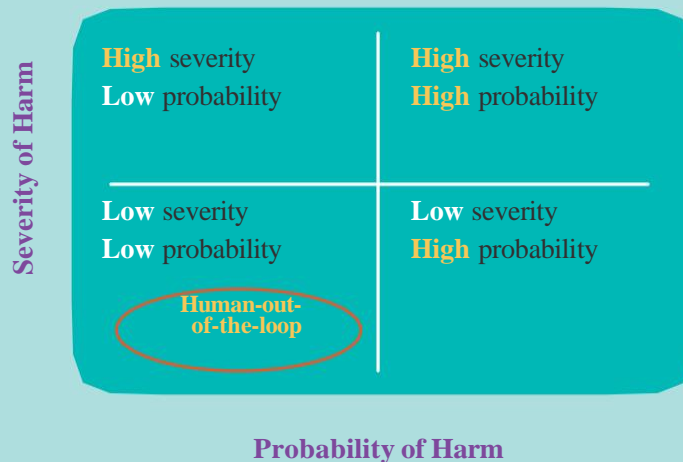


### Магадлал-эрсдлийн үнэлгээ

Хор уршгийн тодорхойлолт гэдэг нь бүтээгдэхүүний зөвлөмжүүдийг гаргахдаа хэрэглэгчдийн хүлээлтэд нийцээгүй байдал юм. Хувь хүмүүст буруу бүтээгдэхүүний зөвлөмж өгсөн тохиолдолд хор уршигийн ноцтой байдал нь бага байж болно, учир нь хувь хүмүүс эцсийн байдлаар худалдан авалт хийх эсэхээ шийддэг. Хор уршгийн магадлал нь хиймэл оюун ухааны шийдлийн үр ашиг, үр нөлөөнөөс хамааран өндөр эсвэл бага байж болно.

### Шийдвэр гаргах үйл явц дахь хүний оролцооны түвшин

Хор уршгийн ноцтой байдал бага байгаа учраас энэ үнэлгээ нь хүний оролцоогүй арга барилыг (өөрөөр хэлбэл, хүнийг оролцуулахгүйгээр) шаарддаг.



### Тогтмол хяналт шал ша шал шалгалт

Байгууллага нь өөрийн арга барилаа (өөрөөр хэлбэл, хүнийг оролцуулахгүйгээр) байнга хянан үзэж, хор уршгийн ноцтой байдал болон магадлалыг дахин үнэлдэг ба нийгмийн хэм хэмжээ, үнэт зүйлсийн өөрчлөлттэй нийцүүлэн шинэчилдэг.

*Тэмдэглэл: Энэ нь тодорхой хэм хэмжээ, үнэт зүйлсийг ашиглан хийсэн энгийн жишээ юм. Байгууллагууд энэхүү арга барилыг илүү төвөгтэй, нарийн ёс зүйн асуудлуудтай харьцуулахад хиймэл оюун ухааны шийдвэр гаргах загварыг тодорхойлох туршилтаар ашиглаж болно.*



# SUADE LABS:

## ХИЙМЭЛ ОЮУН УХААНЫ ШИЙДВЭР ГАРГАХ ҮЙЛ АЖИЛЛАГААНД ХҮНИЙ ОРОЛЦООНЫ ТҮВШНИЙГ ТОГТООХ ТАЛААРХ ЗУРГИЙН ТАЙЛБАР

Suade Labs (“**Suade**”) нь дэлхий даяар үйл ажиллагаа явуулдаг RegTech фирм бөгөөд Дэлхийн эдийн засгийн форумаас технологийн анхдагч гэдгээрээ хүлээн зөвшөөрөгдсөн. Suade нь санхүүгийн байгууллагуудад их хэмжээний нарийвчилсан өгөгдлийг боловсруулах, шаардлагатай зохицуулалтын өгөгдөл, тооцоолол болон тайлангуудыг шаардлагатай хяналт, засаглалтайгаар гаргах боломжийг олгодог хиймэл оюун ухааны шийдлийг санал болгодог. Мөн Suade-ийн шийдэл нь хэрэглэгчдэд одоо мөрдөгдөж буй зохицуулалтын нөлөөллийг, түүнчлэн тухайн хууль тогтоомжийн тус бүрийн нөлөөллийг шинжлэх боломжийг олгодог.

Хиймэл оюун ухааныг ашиглан шийдвэр гаргах үйл явц дахь хүний оролцооны түвшнийг тогтоохдоо Suade дараах гол хүчин зүйлсийг харгалзан үзсэн:

- a. Алгоритмын үр дүнг зөв тайлбарлахад шаардагдах салбарын мэдлэгийн түвшин (жишээ нь: хууль эрх зүйн болон бодлого боловсруулалтын мэдлэг).
- b. Хэрэв хиймэл оюун ухааны хэрэгсэл зохицуулалтын нөлөөллийг зөв шинжилж, зохицуулалтын шаардлагыг хангах зөв зөвлөмж өгч чадахгүй бол, зохицуулалтын шаардлага биелүүлэхгүй байх зардал.

Suade-ийн шийдэл нь хүний мэргэжилтнүүдээс тодорхой хэмжээний салбарын мэдлэг шаарддаг бөгөөд хиймэл оюун ухааны шийдлээс гарсан буруу зөвлөмжийн улмаас зохицуулалтын шаардлагыг биелүүлээгүйгээс үүсэх зардал нь хэрэглэгчдэд ихээхэн ач холбогдолтой байдаг тул Suade нь хиймэл оюун ухааны шийдэлдээ "**human-in-the-loop**" буюу хүний оролцоотой арга барилыг нэвтрүүлсэн.

Нөгөө талаас, хиймэл оюун ухааны загварыг тохируулахад Suade нь "**human-over-the-loop**" буюу хүний хяналттай арга барилыг ашигладаг. Ерөнхийдөө, Suade нь хиймэл оюун ухааны загварыг худал эерэг (false positives) илрүүлэхийг автоматаар илүүд үзэхээр тохируулдаг. Гэсэн хэдий ч, хэрэглэгчдийн судалгаагаар зарим хэрэглэгчид худал сөрөг (false negatives) илрүүлэхийг илүүд үздэг болохыг тодруулсан. Тиймээс, Suade нь хиймэл оюун ухааны загварыг хэрэглэгчдийн ялгаатай хүсэл сонирхолд нийцүүлэхийн тулд "**human-over-the-loop**" арга барилыг ашигладаг бөгөөд ингэснээр алгоритмын үр дүн нь худал эерэг эсвэл худал сөрөг илрүүлэхийг илүүд үзэх эсэхийг хэрэглэгчдийн хүсэлтийн дагуу тохируулах боломжтой болдог.

# GRAB:

## ХИЙМЭЛ ОЮУН УХААНААР ДЭМЖИГДСЭН ШИЙДВЭР ГАРГАХ ҮЙЛ ЯВЦАД ХҮНИЙ ОРОЛЦООНЫ ТҮВШНИЙГ ТОГТООХ ТАЛААРХ ЗУРГИЙН ТАЙЛБАР

Grab бол Сингапурт төвтэй, тээврийн үйлчилгээ, хоол хүргэлт, цахим төлбөрийн шийдлүүдийг санал болгодог компани юм. 'Grab' нь AI-г платформдоо өргөнөөр ашигладаг ба үүнд явах чиглэл хуваарилах, аюулгүй байдлын ослыг илрүүлэх, залилангийн гүйлгээг таних зэрэг орно. Ялангуяа, Grab нь AI-г ашиглан нийт аяллын хуваарилалтын чанарыг сайжруулах болон аяллын цуцлалтыг багасгах зорилгоор ашигладаг.

Аяллыг амжилттай хуваарилахын тулд, Grab-ийн хиймэл оюун ухааны загвар нь жолооч нарын дараах гол хүчин зүйлс дээр үндэслэсэн хүсэл сонирхлыг харгалзан үздэг:

- a. Тодорхой төрлийн аяллын хүсэл сонирхол;
- b. Өдөр тутмын ажлынхаа эхлэх болон дуусах байршлыг хүссэн байршил;
- c. Бусад сонгомол жолоодлогын зан төлөвүүд.

Аяллын хуваарилалтад хиймэл оюун ухаан ашиглан шийдвэр гаргах үйл явц дахь хүний оролцооны түвшнийг тогтоохдоо, Grab дараах гол хүчин зүйлсийг харгалзан үзсэн:

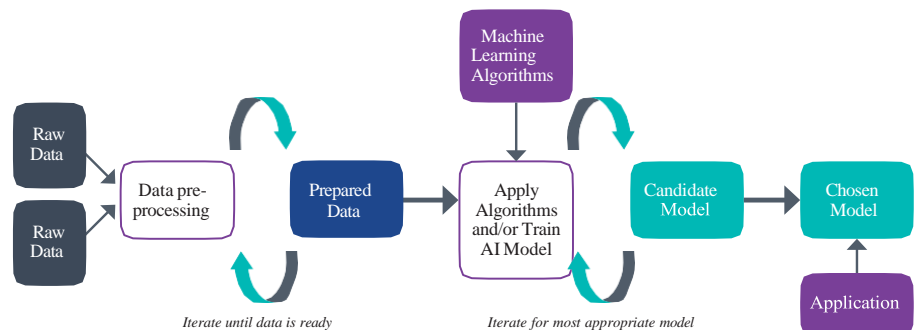
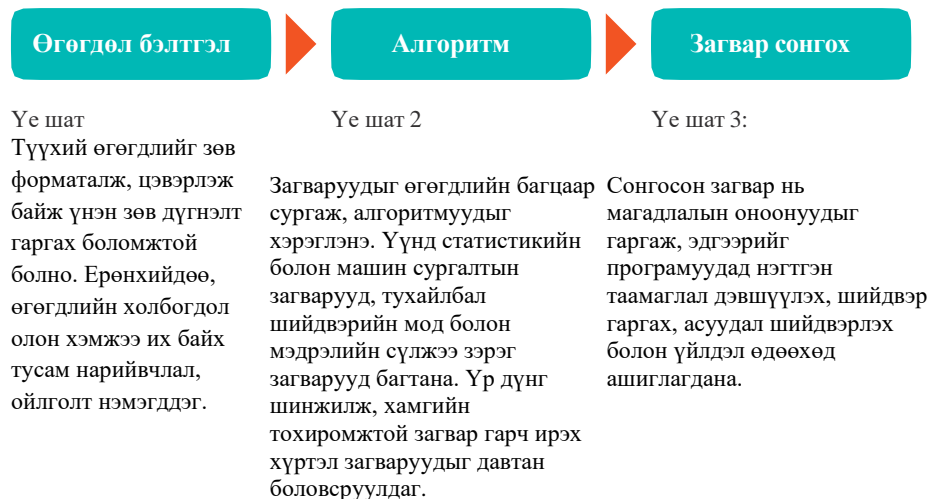
- a. Бодит цагийн шийдвэр гаргалтын цар хүрээ. Grab нь минут тутамд 5000 гаруй аяллыг хуваарилах шаардлагатай байдаг бөгөөд хэрэв хүн бүрийг хянан үзэх шаардлагатай болвол энэ нь үр ашиг, зардлын хувьд үйлчлүүлэгчдэд нөлөөлнө;
- b. Хиймэл оюун ухааны загвар нь оновчгүй ажиллах тохиолдолд хэрэглэгчдэд үзүүлэх хор уршгийн ноцтой байдал болон магадлал.

Бусад хүчин зүйлсийн дунд, Grab нь дараах хоёр хүчин зүйлсийг харгалзан үзсэн: (1) Хүний хувьд богино хугацаанд ийм их хэмжээний аяллыг хуваарилах нь техникийн хувьд боломжгүй; (2) Оновчгүй аяллын хуваарилалт хийх нь амь насанд бага зэрэг эсвэл огт хохирол учруулахгүй. Тиймээс, Grab нь аяллын хуваарилалтад ашиглагдах хиймэл оюун ухааны загвартаа "human-out-of-the-loop" буюу хүний оролцоогүй арга барилыг нэвтрүүлэхээр шийдсэн бөгөөд хиймэл оюун ухааны загварын оновчтой ажиллагааг хангахын тулд байнга хянан шалгаж байдаг.



## ҮЙЛ АЖИЛЛАГААНЫ МЕНЕЖМЕНТ

- 3.19 Энэхүү хэсэг нь байгууллагуудад хиймэл оюун ухаан нэвтрүүлэх үйл явцынхаа үйл ажиллагааны тал дээр хариуцлагатай арга хэмжээ авхад туслах зорилготой. Байгууллагын өгөгдөл, алгоритм, хиймэл оюун ухааны загварын засаглалын сайн практик зөвлөмжийг өгөхийн тулд лавлагаа хиймэл оюун ухаан нэвтрүүлэх үйл явцыг тогтоож өгсөн.
- 3.20 Загвар хүрээ нь байгууллага хиймэл оюун ухааны шийдлийг хэрэгжүүлэх үе шатуудыг тайлбарлахын тулд дараах ерөнхий хиймэл оюун ухааны загвар боловсруулах болон ашиглах үйл явцыг ашигладаг. Энэ үйл явц нь үргэлж нэг чиглэлтэй байдаггүй гэдгийг тэмдэглэх хэрэгтэй – энэ нь тасралтгүй суралцах үйл явц байж болно, ихэвчлэн байдаг.



<sup>5</sup> "Машин сургалтыг өргөн цар хүрээнд хэрэгжүүлэх нь" Microsoft Azure (2018 оны 12-р сарын 2) <https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/machine-learning-at-scale> (2019 оны 12-р сард хандсан).

- 3.21 Хэрэгжүүлэлтийн явцад, шугаман регрессийн алгоритм, шийдвэрийн мод эсвэл мэдрэлийн сүлжээ гэх мэт алгоритмуудыг сургалтын өгөгдлийн сан дээр шинжилгээ хийхэд ашигладаг. Үүний үр дүнд гарсан алгоритмын загваруудыг шалгаж, хэрэглээнд хамгийн тохиромжтой үр дүн гаргадаг загвар бий болтол алгоритмуудыг давтан боловсруулдаг. Энэ загвар болон түүний үр дүнг дараа нь таамаглал гаргах, шийдвэр гаргах, асуудал шийдэх болон үйлдэл эхлүүлэх зэрэгт програмуудад нэгтгэдэг. Өгөгдөл болон алгоритм/загварын нягт холбоо энэ хэсгийн Загварын Хүрээний гол анхааралд байна.

### ЗАГВАР БОЛОВСРУУЛАХ ӨГӨГДӨЛ

- 3.22 Загвар бүтээхэд ашиглагдах өгөгдлийн сан нь олон эх сурвалжаас ирж болох ба хувийн болон хувийн бус өгөгдөл хоёуланг нь агуулж болно. Эдгээр эх сурвалж тус бүрээс авсан өгөгдлийн чанар болон сонголт нь хиймэл оюун ухааны шийдлийн амжилтад чухал ач холбогдолтой. Хэрэв загварыг хазайлттай, буруу эсвэл төлөөлөх чадваргүй өгөгдлийг ашиглан бүтээсэн бол загвараас гарсан санаандгүй ялгаварлан гадуурхах шийдвэрийн эрсдэл нэмэгдэх болно.

*Хиймэл оюун ухааны шийдлийн үр ашигтай байдлыг хангахын тулд, өгөгдлийн чанар, загвар сургалт болон загвар сонголтын хариуцлагатай байгууллагын холбогдох хэлтсүүд хамтран ажиллаж, өгөгдлийн хариуцлагын сайн туршлагыг хэрэгжүүлэх ёстой.*

- 3.23 Загваруудыг сургалтанд хамруулах болон сонгоход оролцож буй хүмүүс нь дотоод ажилтан эсвэл гадаад үйлчилгээ үзүүлэгч байж болно. Ухаалаг системд хэрэгжүүлэх загваруудыг дотоод хэлтсийн эзэнтэй байх нь хамгийн тохиромжтой бөгөөд энэ хүн нь ямар загварыг хэрэгжүүлэх талаар шийдвэр гаргадаг. Хиймэл оюун ухааны шийдлийн үр ашигтай байдлыг хангахын тулд, өгөгдлийн чанар, загвар сургалт болон загвар сонголтын хариуцлагатай байгууллагын холбогдох хэлтсүүд хамтран ажиллаж, өгөгдлийн хариуцлагын сайн туршлагыг хэрэгжүүлэх нь үр дүнтэй байх болно. Үүнд дараах зүйлс багтана:

**а. Өгөгдлийн гарал үүслийг ойлгох:** Энэ нь өгөгдөл анх хаанаас ирсэн, хэрхэн цуглуулсан, зохион байгуулсан болон байгууллагын дотор хэрхэн шилжсэн, мөн цаг хугацааны явцад түүний нарийвчлалыг хэрхэн хадгалсан талаар мэдэхийг хэлнэ. Өгөгдлийн гарал үүслийг харааны хэлбэрээр дүрслэн, өгөгдөл эх сурвалжаас зорилтот газар руугаа хэрхэн шилжиж, замдаа хэрхэн өөрчлөгдөж, бусад өгөгдөлтэй хаана харилцаж, төлөөлөл нь хэрхэн өөрчлөгдөж байгааг хянах боломжтой. Өгөгдлийн гарал үүслийн гурван төрөл байдаг:

- i. Урвуу өгөгдлийн гарал үүсэл: Өгөгдлийг эцсийн хэрэглээнээс нь эхлүүлэн эх сурвалж хүртэлх хугацааны урсгалыг хянах.
- ii. Урагш чиглэсэн өгөгдлийн гарал үүсэл: Өгөгдлийг эх сурвалжаас нь эхлүүлэн эцсийн хэрэглээ хүртэлх урсгалыг хянах.
- iii. Эхлэлээс төгсгөл хүртэлх өгөгдлийн гарал үүсэл: Хоёр аргыг хослуулан, өгөгдлийн эх сурвалжаас эцсийн хэрэглээ хүртэл болон эцсийн хэрэглээнээс эх сурвалж хүртэлх бүх үйл явцыг хянах.

Өгөгдлийн гарал үүслийн бүртгэлийг хадгалах нь байгууллагад өгөгдлийн чанарыг эх үүсвэр болон дараагийн өөрчлөлтүүдийн үндсэн дээр тогтоох, алдааны боломжит эх үүсвэрийг мөрдөх, өгөгдлийг шинэчлэх, мөн өгөгдлийг эх үүсвэрт нь холбон тодорхойлоход тусалдаг.

Зарим тохиолдолд, өгөгдлийн эх үүсвэрийг тогтоох нь хэцүү байж болох юм. Жишээ нь, итгэмжлэгдсэн гуравдагч этгээдээс авсан өгөгдлийн сангууд нь олон эх сурвалжаас нэгтгэсэн өгөгдлийг агуулж байж болно. Ийм өгөгдлийг ашиглах эрсдлийг үнэлж, зохих ёсоор удирдах нь байгууллагуудад зөв зохистой юм.

- b. **Өгөгдлийн чанарыг хангах:** Байгууллагууд өгөгдлийн чанарт нөлөөлж болзошгүй хүчин зүйлсийг ойлгож, шийдвэрлэхийг уриалж байна. Үүнд дараах зүйлс орно:
- i. Өгөгдлийн сангийн нарийвчлал, энэ нь өгөгдлийн санд байгаа утгууд нь тайлбарлагдсан нэгжүүдийн жинхэнэ шинж чанарт хэр сайн нийцэж байгааг илэрхийлнэ;
  - ii. Өгөгдлийн сангийн бүрэн бүтэн байдал, энэ нь шинж чанар болон зүйлсийн аль алиных нь хувьд;
  - iii. Өгөгдлийн сангийн үнэн зөв байдал, энэ нь өгөгдлийн найдвартай эх сурвалжаас ирсэн эсэх зэрэг өгөгдлийн хэр найдвартайг илэрхийлнэ;
  - iv. Өгөгдлийн санг хамгийн сүүлд хэзээ цуглуулсан эсвэл шинэчилсэн;
  - v. Өгөгдлийн сангийн ач холбогдол болон өгөгдөл цуглуулахтай холбоотой нөхцөл байдал, энэ нь өгөгдлийг зориулагдсан зорилгоор тайлбарлах болон найдах байдлыг нөлөөлж болзошгүй;
  - vi. Олон өгөгдлийн сангаас нийлүүлсэн өгөгдлийн сангийн бүрэн бүтэн байдал, энэ нь олборлолт болон өөрчлөлт хэр сайн хийгдсэн эсэхийг илэрхийлнэ;
  - vii. Өгөгдлийн сангийн ашиглалт, өгөгдлийн сан нь машинд ойлгомжтой хэлбэрээр хэр сайн бүтэцлэгдсэн байгааг;
  - viii. Хүний оролцоо (жишээ нь, хүний шүүлтүүр, шошго хүлээлгэх эсвэл өгөгдлийг засварласан эсэх).
- c. **Дотоод хазайлтыг багасгах:** Хиймэл оюун ухааныг хамаарах олон төрлийн хазайлт байдаг. Загварын Хүрээ нь өгөгдлийн сан дахь дотоод хазайлтыг анхаарч үздэг бөгөөд энэ нь хүсээгүй үр дүн, тухайлбал, санаандгүйгээр ялгаварлан гадуурхах шийдвэр гаргах зэрэгт хүргэж болзошгүй. Байгууллагууд өөрсдийн хиймэл оюун ухааны системд өгсөн өгөгдөл нь дотоод хазайлтыг агуулж болзошгүйг анхаарч, ийм хазайлтыг багасгах алхмуудыг хийхийг уриалж байна. Өгөгдөлд түгээмэл тохиолддог хоёр төрлийн хазайлт орно:

- i. **Сонголтын хазайлт:** Энэ хазайлт нь загвар бүтээхэд ашиглагдаж буй өгөгдөл нь загварын хүлээн авах эсвэл үйл ажиллагаа явуулах бодит өгөгдөл эсвэл орчныг бүрэн төлөөлж чадахгүй үед үүсдэг. Өгөгдлийн сан дахь сонголтын хазайлтын түгээмэл жишээнүүд нь орхигдсон хазайлт болон хэвшмэл хазайлт юм. Орхигдсон хазайлт нь өгөгдлийн сангаас тодорхой шинж чанаруудыг орхихыг тайлбарладаг. Жишээлбэл, зөвхөн Азийн нүүр царайг агуулсан өгөгдлийн санг нүүр царай таних сургалтанд ашиглахад, хэрэв тухайн хүн ам нь Ази бус хүмүүсийг ч агуулж байвал энэ нь орхигдсон хазайлтыг үүсгэнэ. Ажлын өдөр төв бизнес дүүрэгт байгаа тээврийн хэрэгслийн төрлүүдийн өгөгдлийн санг Сингапурт боломжтой тээврийн төрлүүдийг загварчлахад ашиглахад, энэ нь зөвхөн автомашин, автобус, мотоциклд төвлөрч, дугуйг дутуу төлөөлж байгаа тохиолдолд хэвшмэл хазайлтыг харуулж болно.
- ii. **Хэмжилтийн хазайлт:** Энэ хазайлт нь өгөгдөл цуглуулах төхөөрөмж нь өгөгдлийг системтэйгээр тодорхой чиглэлд хазайлгахад хүргэдэг үед үүсдэг. Жишээлбэл, сургалтын өгөгдлийг өнгөт шүүлтүүрийг унтраасан камер ашиглан авсан бол, энэ нь машин сургалтын үр дүнд хазайлт үүсгэнэ.

Өгөгдлийн санд дотоод хазайлтыг илрүүлэх, шийдвэрлэх нь хялбар биш байж болох ч, байгууллагууд нь олон төрлийн өгөгдлийн сан (өөрөөр хэлбэл, олон янзын найдвартай эх сурвалжаас өгөгдөл цуглуулах) ашиглан дотоод хазайлтын эрсдэлийг багасгах боломжтой. Өөр нэг арга нь өгөгдлийн санг өгөгдлийн шинж чанар болон өгөгдлийн зүйлсийн хувьд аль болох бүрэн бүтэн байлгах явдал юм. Өгөгдлийн шинж чанаруудыг дутуу хасах нь дотоод хазайлт илрүүлэх, шийдвэрлэхэд хүндрэл учруулж болно.



- d. **Сургалт, туршилт, баталгаажуулалтад зориулсан өөр өөр өгөгдлийн сан:** Сургалт, туршилт, баталгаажуулалтад зориулсан өөр өөр өгөгдлийн сан шаардлагатай. Загварыг сургалтын өгөгдлөөр сургалтанд оруулж, загварын нарийвчлалыг туршилтын өгөгдлөөр тодорхойлдог. Хэрэв боломжтой бол, загварыг системтэй хазайлттай эсэхийг шалгахын тулд өөр өөр хүн ам зүйн бүлгүүд дээр туршиж үзэх замаар аливаа бүлэг системтэйгээр давуу буюу сул талтай болж байгаа эсэхийг ажиглаж болно.

Эцэст нь, сургалтанд хамрагдсан загварыг баталгаажуулах өгөгдлийн сан ашиглан баталгаажуулж болно. Их хэмжээний өгөгдлийн санг эдгээр зорилгоор дэд хэсгүүдэд хуваах нь өгөгдлийн нарийвчлал болон төлөөлөх чадварыг бууруулахгүй бол сайн практик гэж үздэг. Гэхдээ энэ нь боломжгүй үед (жишээ нь: байгууллага нь их хэмжээний өгөгдлийн сантай ажиллахгүй эсвэл дамжуулалтын сургалтын үед урьдчилан сургасан загваруудыг ашиглаж байгаа тохиолдолд), байгууллагууд системтэй хазайлтын эрсдлийг ойлгож, зохих хамгаалалтыг хэрэгжүүлэхийг уриалж байна.

- e. **Өгөгдлийн санг тогтмол хянах болон шинэчлэх:** Өгөгдлийн санг (сургалт, туршилт, баталгаажуулалтын өгөгдлийн санг оролцуулан) тогтмол хянаж, нарийвчлал, чанар, шинэ байдал, ач холбогдол, найдвартай байдлыг хангах нь зохистой байдаг. Шаардлагатай тохиолдолд, өгөгдлийн санг үйлдвэрлэлд ашиглагдаж буй хиймэл оюун ухааны загваруудаас олж авсан шинэ оруулах өгөгдлөөр шинэчлэх боломжтой. Ийм шинэ оруулах өгөгдлийг ашиглах үед, нэгэнт нэг удаа загвараар дамжсан шинэ оруулах өгөгдөл ашиглах нь бэхжүүлэх хазайлт үүсгэж болзошгүй тул байгууллагууд болгоомжтой байх хэрэгтэй.

- 3.24 Хиймэл оюун ухааны загваруудыг сургалтанд ашиглах үед зөвхөн хувийн бус өгөгдөл (үүнд нэргүй болгосон хувийн өгөгдөл орно) ашигласан ч гэсэн дээр дурдсан өгөгдлийн хариуцлагын сайн практикууд нь чухал хэвээр байна.

# SUADE ЛАБОРАТОРИ:

## ЗАГВАР ХӨГЖҮҮЛЭХЭД ӨГӨГДӨЛИЙГ УДИРДАХ ТАЛААРХ ЗУРАГЛАЛ

Suade (дээр танилцуулсан) нь санхүүгийн байгууллагуудад өөрсдийн үйл ажиллагаа явуулдаг хууль эрх зүйн шаардлагад нийцсэн шаардлагатай өгөгдөл, тайланг үүсгэхэд тусалдаг хиймэл оюун ухаанаар идэвхжүүлсэн шийдлийг боловсруулсан.

Suade-ийн хиймэл оюун ухааны загвар хөгжүүлэхэд ашигладаг өгөгдөл нь чанар, гүйцэтгэлд шууд нөлөөлдөг тул Suade хэд хэдэн сайн өгөгдөл хариуцлагын практикийг нэвтрүүлсэн. Жишээлбэл, хууль эрх зүйн өгөгдлийг найдвартай, баталгаатай эх сурвалжаас авсан эсэхийг баталгаажуулахын тулд Suade зөвхөн холбогдох зохицуулагчдаас хууль эрх зүйн өгөгдлийг авч шинэчилдэг. Мөн Suade ашигласан өгөгдлийн багцуудыг нэмэлт мета өгөгдлөөр тэмдэглэдэг. Энэ нь шаардлагатай үед, жишээлбэл, зөрүү олдсон үед өгөгдлийн багцуудыг анхны эх сурвалж руу нь эргүүлэн хайхад тусалдаг. Цаашлаад, хиймэл оюун ухааны загварт ямар өгөгдлийн багцууд ашиглагдсан болохыг олж мэдэхийн тулд Suade загвар хөгжүүлэлтийн холбогдох мэдээллийг баримтжуулж, өгөгдлийн санд хадгалдаг.

Suade мөн хариуцлагатай өгөгдөл тэмдэглэгээгээр дамжуулан хиймэл оюун ухааны загваруудын дотоод эрсдэлийг багасгадаг. Өгөгдлийг тэмдэглэдэг хүмүүсийн (өгөгдөл тэмдэглэгээчид) тоог нэмэгдүүлэх замаар Suade хиймэл оюун ухааны загваруудын гаралтыг аль болох төвийг сахисан байлгахыг зорьж, өгөгдлийг тэмдэглэхдээ өгөгдлийн орчноос (ихэвчлэн текстээс бүрддэг) нөлөөлөх эрсдэлийг багасгадаг. Өөрөөр хэлбэл, өгөгдлийг тэмдэглэхэд практикт байгаа олон хүмүүсийг ашигласнаар тэмдэглэгчийн хазайлтын эрсдэлийг багасгадаг.

Мөн Suade өгөгдлийг тэмдэглэхэд хялбар болгохын тулд тэмдэглэгээний системийг боловсруулсан. Энэ системийг алгоритмд ашиглах сургалтын өгөгдлийг үүсгэхэд ашигладаг. Suade олон тэмдэглэгээчдийг удирдах чадвараа сайжруулах болон загвар сургалтанд ашиглах өгөгдлийн багцыг илүү сайн сонгохын тулд энэхүү тэмдэглэгээний системийг цаашид хөгжүүлэх болно. Suade мөн тэмдэглэгээний системийг шинэ өгөгдлөөр тогтмол шинэчилдэг. Шинэ сургалтын өгөгдөл нь хиймэл оюун ухааны загварт давтан оруулдаг. Энэ байдлаар хиймэл оюун ухааны загвар шинэ өгөгдлийн багцаас тасралтгүй суралцах боломжтой.

Suade өгөгдөл хариуцлагын өөр нэг практик нь өгөгдлийн хувиргалтын янз бүрийн үе шатанд баталгаажуулах схемийн шалгалтыг ашиглах явдал юм. Энэ нь Suade өгөгдлийн схем нь эх сурвалжаас өгөгдлийг үнэн зөвөөр төлөөлж байгаа эсэхийг баталгаажуулдаг үйл явц бөгөөд өгөгдлийн формат, агуулга зэрэг хүчин зүйлүүдэд алдаа гардаггүйг хангах зорилготой юм.

# ПАЙМЕТРИКС:

## ЗАГВАР ХӨГЖҮҮЛЭХЭД ӨГӨГДЛИЙН ТӨВИЙГ САХИХТАЙ ХОЛБООТОЙ ЗУРАГЛАЛ

Пайметрикс нь нейробиологийн мэдлэг болон аудит хийгдсэн хиймэл оюун ухааны загварыг ашиглан ажил горилогчдыг илүү урьдчилан таамаглах чадвартай, бага хазайлтаар үнэлэхэд тусалдаг технологийн үйлчилгээ үзүүлэгч юм. Хиймэл оюун ухааны загвар хөгжүүлэхийн тулд Пайметрикс дараах алхмуудыг хийдэг:

- Үйлчлүүлэгчдийнхээ шилдэг гүйцэтгэлтэй ажилчдыг Пайметриксийн үнэлгээнд хамруулж, тухайн ажлын байрны шаардлагад хамгийн сайн нийцэх ажилтны шинж чанарын профайлыг бий болгодог;
- Шинж чанарын профайлыг үйлчлүүлэгчийн хүний нөөцийн багтай баталгаажуулдаг;
- Пайметриксийн тоглоомжуулсан үнэлгээгээр дамжуулан ажил горилогчдын зан үйлийн өгөгдлийг цуглуулж, шинж чанарын профайлын үндсэн дээр ажил горилогчдын тохиромжтой байдлыг үнэлдэг.

Нийгмийн мэдрэмтгий шинж чанаруудтай харьцах, хиймэл оюун ухааны загварт ашигладаг өгөгдөлд агуулагдах суурь болон санаандгүй хазайлтын эрсдэлийг бууруулахын тулд Пайметрикс дараах арга хэмжээнүүдийг авдаг:

- Тогтсон нейробиологийн судалгаанд үндэслэсэн объектив өгөгдлийг ашигладаг (жишээ нь, нарийн ширийнийг анхаарах, анхаарал төвлөрүүлэх чадвар болон ой санамжийн чадвар), эдгээр нь ерөнхийдөө хүйс, арьс өнгө болон насны бүлгүүдийн хувьд тогтвортой байдаг.
- Бүх хиймэл оюун ухааны загваруудыг идэвхтэйгээр хазайлтаас чөлөөлж, тухайн загварууд ажиллах орчныг бүрэн төлөөлөхийг баталгаажуулдаг, ингэснээр хиймэл оюун ухааны загварууд хүн ам зүйн шинж чанараас нь шалтгаалан хүмүүст хохирол учруулахгүй байх нөхцөлийг хангана:
  - » Шударга байдлын стандартыг хууль эрх зүйн шаардлагууд тодорхойлдог. Урьдчилсан ажилд авах үнэлгээний хувьд, Пайметриксийн загварууд “дөрөвний тавын дүрэм” гэж нэрлэгддэг шалгалтыг давж гарах ёстой бөгөөд энэ нь ажил эрхлэлийн хуульд ихэвчлэн дурдагддаг. АНУ-ын Тэг Үйл Ажиллах Боломжийн Хорооны (“ЕЕОС”) дагуу, хуулиар хамгаалагдсан аливаа бүлгийн сонгон шалгаруулалтын түвшин нь олонхийн бүлгийн сонгон шалгаруулалтын түвшний дор хаяж 80%-ийг бүрдүүлэх ёстой. Жишээ нь, ажил олгогч 200 тохирох горилогчийг (100 эрэгтэй, 100 эмэгтэй) шалгаруулж байгаа бол, 50 эрэгтэйг сонгож буй загвар дор хаяж 40 эмэгтэйг ч мөн сонгох ёстой.
  - » Пайметрикс нь хиймэл оюун ухааны загварыг төрөл бүрийн хүн ам зүйн бүлгүүдийн хэрэглэгчдийн өгөгдөл дээр туршиж, өгөгдлийн санамсаргүй хээ загвараар суралцахгүй байхыг баталгаажуулж, боломжит хазайлтын эрсдлийг шийдвэрлэнэ

- »Пайметрикс газарзүйн ач холбогдол эсвэл хууль эрх зүйн шаардлагад үндэслэн нэмэлт хүн ам зүйн мэдээллийг хазайлтаас цаашид чөлөөлөх ажлыг гүйцэтгэнэ. Пайметрикс хазайлтын харьцааг ашиглан, хүн ам зүйн ангилал бүрийн (жишээ нь, хүйс болон угсаатны бүлэг) хамгийн өндөр амжилт үзүүлсэн бүлгийн хувь тэнцүүлсэн амжилтын түвшинг хамгийн бага амжилт үзүүлсэн бүлэгтэй харьцуулдаг.
- » Хиймэл оюун ухааны загварыг зөвхөн ЕЕОС-ийн стандартыг хангасан тохиолдолд л хэрэгжүүлэх болно.
- Хиймэл оюун ухааныг хэрэгжүүлсний дараа Пайметрикс:
  - » Хиймэл оюун ухааны загварын шийдвэрийг бодит ажил горилогчдод үзүүлэх сөрөг нөлөөллийг шалгах;
  - » Системийн урт хугацааны таамаглалуудын тухайн ажлын байран дахь хадгалалтын нөлөөллийг дахин шалгана.

Хэрэв хазайлт илэрсэн тохиолдолд, хэрэгжүүлэхээс өмнө болон дараа нь Пайметрикс хиймэл оюун ухааны загварыг ажил горилогчдын хувьд шударга байдлыг хангахын тулд өөрчлөлт хийн, хиймэл оюун ухааны загварын урьдчилан таамаглах чадварыг хадгалах болно.

## АЛГОРИТМ БА ЗАГВАР

3.25 Хиймэл оюун ухааны системүүд нь хиймэл оюун ухааны загвар дахь алгоритмаар идэвхжүүлэгддэг олон шинж чанар эсвэл үйл ажиллагаатай байж болно. Алгоритмуудын ил тод байдлыг сайжруулахын тулд тайлбарлах боломжтой байдал, давтагдах чадвар, бат бөх байдал, тогтмол тааруулах, давтагдах чадвар, мөрдөх чадвар болон аудит хийх чадвар зэрэг арга хэмжээнүүдийг хэрэгжүүлж болно. Эдгээр арга хэмжээнүүдийн аль чухалыг бүх алгоритмууд дээр хэрэгжүүлэх нь практикт боломжгүй эсвэл зардал ихтэй байж болох юм.

Байгууллагуудыг эрсдэлд суурилсан хандлагыг авч, хоёр талт үнэлгээ хийхийг уриалж байна. Эхлээд, эдгээр арга хэмжээнүүд хамааралтай болохыг харуулдаг шинж чанар эсвэл үйл ажиллагааны дэд хэсгийг тодорхойлох хэрэгтэй. Хоёрдугаарт, эдгээр арга хэмжээнүүдээс аль нь харилцагчдынхаа итгэлийг бий болгоход хамгийн үр дүнтэй болохыг тодорхойлох хэрэгтэй. Зарим арга хэмжээнүүд, жишээ нь, тайлбарлах боломжтой байдал (эсвэл давтагдах чадвар, тайлбарлахад хялбар биш загваруудыг ашиглах үед), бат бөх байдал болон тогтмол тааруулах нь хангалттай чухал бөгөөд байгууллагын хиймэл оюун ухааны хэрэгжүүлэлтийн үйл явцад тодорхой хэмжээгээр оруулах боломжтой. Бусад арга хэмжээнүүд, жишээ нь, давтагдах чадвар, мөрдөх чадвар болон аудит хийх чадвар нь илүү их нөөц шаарддаг бөгөөд тусгай шинж чанарууд эсвэл тодорхой нөхцөл байдалд хамааралтай байж болно.

## Тайлбарлах боломжтой байдал

3.26 Тайлбарлах боломжтой байдлыг, хэрэгжүүлсэн хиймэл оюун ухааны загварын алгоритмууд хэрхэн ажилладаг болон/эсвэл шийдвэр гаргах үйл явц нь загварын таамаглалыг хэрхэн тусгадаг талаар тайлбарласнаар олж авдаг. Хиймэл оюун ухааны гаргасан таамаглалыг тайлбарлах чадвар нь ойлголт болон итгэлийг бий болгоход оршдог. Хиймэл оюун ухааны шийдэлд ашигласан алгоритмын үйл ажиллагааг болон тодорхой таамаглалд хэрхэн хүрч байгааг тайлбарлаж чадаж байгаа тохиолдолд уг алгоритмыг тайлбарлах боломжтой гэж хэлнэ. Алгоритмыг тайлбарлаж чадахгүй үед ч гэсэн, таамаглалууд шийдвэр гаргах үйл явцад хэрхэн үүрэг гүйцэтгэдгийг тайлбарлах замаар ойлголт болон итгэлийг бий болгож болно.

3.27 Хиймэл оюун ухааны шийдэл хэрэгжүүлж буй байгууллагуудад дараах практикуудыг хэрэгжүүлэхийг зөвлөж байна:

а. Загвар сургалт ба сонголт нь ухаалаг систем (өөрөөр хэлбэл, хиймэл оюун ухааны технологийг агуулсан систем) хөгжүүлэхэд зайлшгүй шаардлагатай. Загварын сургалт болон сонголтын үйл явцууд хэрхэн явагдсаныг, гаргасан шийдвэрүүдийн шалтгаануудыг болон илэрсэн эрсдлүүдийг шийдвэрлэх арга хэмжээнүүдийг баримтжуулснаар байгууллага нь дараах шийдвэрүүдийн талаар тайлагнах боломжтой болно. Энэ хүрээнд, Автоматжуулсан Машин Сургалтын салбар нь машин сургалтын ажлын урсгалын ихэнх хэсгийг автоматжуулахыг зорьдог. Үүнд шинж чанар боловсруулах, шинж чанар сонгох, загвар сонгох болон гипер параметр тохируулах зэрэг үйл явц орно. Ийм төрлийн хэрэгслүүдийг ашиглаж буй байгууллагууд нь автоматжуулсан машин сургалтын хандлага болон сонгогдсон загваруудын ил тод байдал, тайлбарлах боломжтой байдал, мөрдөх чадварыг харгалзан үзэж болно.

б. Шийдлүүдийн загвар болон хүлээгдэж буй зан төлвийг бүтээгдэхүүн эсвэл үйлчилгээний тайлбар болон системийн техникийн тодорхойлолтын баримт бичигт тусгаж оруулах нь хувь хүн болон/эсвэл зохицуулагчдад хариуцлага хүлээлгэхийг харуулдаг. Энэ нь мөн тодорхой шинж чанар, атрибут эсвэл загваруудыг яагаад бусдаас илүү сонгосон тухай загварын шийдвэрүүдийг багтааж болно. Эдгээр алхмууд нь хиймэл оюун ухааны загварыг илүү тодорхой болгоход тусалж, загвар хэрхэн ажилладаг талаар ойлгомжтой, хүлээж авахад хялбар мэдээлэл өгөх боломжтой.

Хэрэв байгууллагын хиймэл оюун ухааны систем нь гуравдагч талын хиймэл оюун ухааны шийдэл нийлүүлэгчээс авсан эсвэл худалдан авсан бол байгууллага нь хиймэл оюун ухааны шийдэл нийлүүлэгчээс тусламж хүсэхийг харгалзан үзэж болно. Учир нь тэд шийдэл хэрхэн ажилладаг талаар илүү сайн тайлбар өгөх чадвартай байж магадгүй юм.

с. Туслах тайлбарлах хэрэгслүүд нь хиймэл оюун ухааны загваруудыг, ялангуяа бага тайлбарлагддаг (өөрөөр хэлбэл “хар хайрцаг” системүүд) загваруудыг тайлбарлахад тустай байдаг. Эдгээр хэрэгслүүд нь хиймэл оюун ухааны системийн гаргалгааны үндсэн шалтгааныг системийг ашиглагчдад илүү тайлбарлах боломжтой, ойлгомжтой болгодог. Хиймэл оюун ухааны загварын шийдвэрийн тайлбарлах боломжийг сайжруулахын тулд эдгээр хэрэгслүүдийг хослуулан ашиглах боломжтой.

- 3.28 Техникийн тайлбарлал байнга ойлгомжтой байх албагүй, ялангуяа энгийн хүмүүст. Хиймэл оюун ухааны загваруудын алгоритмууд хэрхэн ажилладгийг шууд тайлбарлахаас илүүтэйгээр далд тайлбарууд илүү ашигтай байж болно. Жишээлбэл, тухайн хүнд эсрэг нөхцөл байдлуудыг (жишээ нь, “таны дундаж өр 15% бага байсан бол таны хүсэлт батлагдах байсан”) болон/эсвэл харьцуулалтуудыг (жишээ нь, “таны профайлтай төстэй профайлтай хэрэглэгчид ижил шийдвэрийг авсан”) өгөх нь байгууллагуудын авч үзэж болох хүчирхэг тайлбарын төрлүүдийн нэг юм.
- 3.29 Гэсэн хэдий ч, зарим тохиолдолд алгоритмуудтай холбоотой мэдээллийг өгөх нь практик биш эсвэл зохимжгүй байж болно. Жишээлбэл, мөнгө угаахтай тэмцэх, мэдээллийн аюулгүй байдал болон луйврын илрүүлэлтэд ашиглагдаж буй алгоритмуудыг задруулах нь муу санаатнуудыг илрүүлэхгүй өнгөрөөх боломжийг олгож болзошгүй; мөн, өмчийн алгоритмуудын талаар эсвэл алгоритмуудын гаргасан шийдвэрүүдийн талаар дэлгэрэнгүй мэдээлэл өгөх нь нууц бизнесийн мэдээллийг ил гаргаж болзошгүй.

<sup>6</sup> Эдгээр хэрэгслүүдийг "туслах" гэж нэрлэдэг, учир нь одоогоор хиймэл оюун ухааны загваруудыг тайлбарлах цогц техникийн шийдэл байхгүй байна. Иймд эдгээр хэрэгслүүд нь хиймэл оюун ухааны загварын ажиллагааны зарим түвшний тайлбарлах боломжийг олгож, туслах үүрэг гүйцэтгэдэг. Эдгээр туслах хэрэгслүүдийн жишээнд орлуулсан загваруудыг ашиглах, хэсэгчилсэн хамааралтын графикууд, глобал хувьсагчийн ач холбогдол/харилцан үйлчлэл, мэдрэмжийн шинжилгээ, эсрэг нөхцөл байдлын тайлбарууд, өөрийгөө тайлбарлах болон анхааралд суурилсан системүүд орно.

## Давтагдах чадвар

3.30 Технологийн хөгжлийн түвшинг харгалзан үзэхэд тайлбарлах боломжтой байдал практик хувьд хүрэх боломжгүй тохиолдолд, байгууллагууд хиймэл оюун ухааны загвараас гарсан үр дүнгийн давтагдах чадварыг баримтжуулахыг авч үзэж болно. Давтагдах чадвар нь ижил нөхцөл байдалд үйлдэл гүйцэтгэх эсвэл шийдвэр гаргах чадварыг тогтмол хангахыг хэлнэ. Үр дүнгийн давтагдах чадвар (алгоритмын тайлбарлах боломжтой байдалтай тэнцэхүйц биш ч) нь гүйцэтгэлийн тогтвортой байдлыг баталгаажуулж, хиймэл оюун ухааны хэрэглэгчдэд илүү итгэл төрүүлэх боломжийг олгодог. Туслах практикууд нь:

- a. **Коммерцийн хэрэгжүүлэлтүүдийн давтагдах чадварыг үнэлэх:** Шууд орчинд хийсэн хэрэгжүүлэлтүүд нь давтагдах чадвартай эсэхийг баталгаажуулахын тулд давтагдах чадварын үнэлгээг хийх.
- b. **Эсрэг нөхцөл байдлын шударга байдлын тестийг гүйцэтгэх:** Эсрэг нөхцөл байдлын шударга байдлын тест нь загварын шийдвэрүүд бодит ертөнц болон арьс өнгө, хүйс гэх мэт мэдрэмтгий атрибутууд өөрчлөгдсөн эсрэг нөхцөл байдлын ертөнцөд ижил байхыг баталгаажуулдаг;<sup>7</sup>
- c. Шийдвэр давтагдах боломжгүй үед онцгой тохиолдлуудыг хэрхэн тодорхойлж, шийдвэрлэх боломжтойг үнэлэх, жишээлбэл, санамсаргүй байдлыг загвараар нэвтрүүлсэн үед;
- d. Онцгой тохиолдлуудыг байгууллагын бодлоготой нийцүүлэхийг баталгаажуулах; Үүнтэй холбогдуулан өгөгдсөн багц баримтууд нь өмнө нь авч үзээгүй шинэ хувьсагчдыг агуулж байгааг таних чадвартай, эдгээр шинэ хувьсагчдыг хүнд онцлон харуулах чадвартай хиймэл оюун ухааны загваруудыг ашиглах нь тустай байж болох юм;
- e. Цаг хугацааны мэдрэмтгий өгөгдөл дээр сургагдсан загварууд хамааралтай хэвээр байхын тулд цаг хугацааны өөрчлөлтийг тодорхойлж, бүртгэх.

<sup>7</sup> James Manyika, Jake Sitberg, and Brittany Presten, "What Do We Do About the Biases in AI?" Harvard Business Review (25 October 2019) <<https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>> (accessed 31 October 2019).



### Бат бөх байдал

- 3.31 Бат бөх байдал гэдэг нь компьютерийн системийн алдаа гаргах үед эсвэл буруу оролт хийх үед гүйцэтгэх чадвар болон буруу оролтын эсрэг зөв ажиллах чадварыг хэлнэ. Загваруудыг хангалттай бат бөх байлгах нь хиймэл оюун ухааны системд итгэх итгэлийг нэмэгдүүлнэ.
- 3.32 Бат бөх байдлын ойлголт нь загваруудын үйлдлийн бүх урьдчилсан нөхцөл ба үр дагаврыг тооцоолох боломжгүйгээс үүсдэг. Загварууд бага хэмжээний оролтын өгөгдөл өөрчлөгдсөнөөр мэдрэгдэхгүй үр дүн гаргаж болно. Бат бөх байдлыг тестлэхийн тулд тохиолдлын суурьтай<sup>8</sup> алдаатай оролтыг шалгаж, эсрэг нөхцөл байдлын тестийг хийж, олон төрлийн хүлээгдээгүй оролтын хувьсагчдыг зохицуулах чадварыг хангахыг зөвлөж байна.
- 3.33 Аль ч загвар төгс бат бөх байж чадахгүй, учир нь оролтын өгөгдлийн бүх өөрчлөлтийг илрүүлэх боломжгүй юм. Тиймээс тасралтгүй суралцах загваруудыг ашиглахдаа, тэдгээр нь урьдчилан таамаглах боломжгүй байдлаар ажиллаж болзошгүй учраас эрсдлүүдийг мэдэж байх хэрэгтэй.

<sup>8</sup> Энэ нь хэрэглэгчийн хүлээн зөвшөөрөх тест (UAT)-ээс ялгаатай юм. UAT нь бодит програм хангамжийн хэрэглэгчид шинээр хөгжүүлсэн програм хангамжийг зах зээлд гаргахаас өмнө бодит ертөнцийн нөхцөлд шаардлагатай даалгавруудыг гүйцэтгэж чадах эсэхийг шалгах үйл явц юм. UAT нь ихэвчлэн програм хангамжийг зах зээлд гаргахын өмнөх чухал алхам болдог.

### Тогтмол тааруулах

- 3.34 Дотоод бодлого болон үйл явцыг **тогтмол загвар тааруулахад** чиглүүлж, цаг хугацааны явцад хэрэглэгчдийн зан төлөвийн өөрчлөлтөд нийцүүлэх нь үр дүнтэй арга юм. Энэ нь байгууллагуудад шинэ оролтын өгөгдлийг агуулсан шинэчилсэн сургалтын өгөгдлийн багц дээр үндэслэн загваруудыг шинэчлэх боломжийг олгодог. Арилжааны зорилго, эрсдэл эсвэл компанийн үнэт зүйлс өөрчлөгдсөн үед загвар тааруулалт шаардлагатай байж болно.
- 3.35 Шалгалт нь төлөвлөсөн үйлдвэрлэлийн орчны динамик байдлыг аль болох тусгах ёстой. Аюулгүй байдлыг хангахын тулд шалгалт нь хиймэл оюун ухааны шийдэл ерөнхийдөө хэр сайн ажиллаж байгааг, мөн алдаа гаргах үед хэрхэн зохицуулж байгааг үнэлэх шаардлагатай байж болно. Жишээлбэл, агуулахын робот саад бэрхшээлүүдийг тойрон ажиллах үүрэгтэй бол янз бүрийн саад бэрхшээл, дотоод орчны өөрчлөлтүүдээр (жишээ нь, өөр өөр өнгийн цамц өмссөн ажилчид) шалгах хэрэгтэй. Хэрвээ үүнийг хийхгүй бол загварууд нь бодит үйлдвэрлэлийн орчны нөхцлийг тусгахгүй байж магадгүй. Хиймэл оюун ухааны загваруудыг бодит орчинд хэрэгжүүлсний дараа идэвхтэй хяналт, шалгалт, тааруулалт хийхийг зөвлөж байна.

### Мөрдөх чадвар

- 3.36 Хиймэл оюун ухааны загвар нь (a) түүний гаргасан шийдвэрүүд, (b) хиймэл оюун ухааны загварын шийдвэрийг гаргах процесс болон өгөгдлийн багцууд (өгөгдөл цуглуулах, өгөгдөл тэмдэглэх болон ашигласан алгоритмууд) ойлгомжтой байдлаар баримтжуулсан бол мөрдөх чадвартай гэж үздэг. Эхнийх нь хиймэл оюун ухаанаар дэмжигдсэн шийдвэрүүдийн мөрдөх чадварыг, харин хоёр дахь нь загварын сургалтын мөрдөх чадварыг илэрхийлнэ.

- 3.37 Байгууллагуудын мөрдөж болохуйц байдлыг дэмжих талаар авч үзэж болох практикт дараахь зүйлс орно.
- a. **Аудитын мөр гаргах:** Загварын сургалт болон хиймэл оюун ухаанаар дэмжигдсэн шийдвэрийг баримтжуулах.
  - b. **Хар хайрцаг бичигч хэрэгжүүлэх:** Бүх оролтын өгөгдлийн урсгалыг барьж авах хар хайрцаг бичигчийг хэрэгжүүлэх. Жишээлбэл, өөрөө жолооддог машинд хар хайрцаг бичигч нь тээврийн хэрэгслийн байрлалыг хянаж, өөрөө жолоодлогын систем хэзээ, хаана хяналтыг авсан, техникийн асуудал гарсан эсвэл жолоочид хяналтыг авахыг хүссэн үед бичиж авдаг.<sup>9</sup>
  - c. **Мөрдөх чадвартай холбоотой өгөгдлийг зөв хадгалах:** Өгөгдлийн доройтол, өөрчлөлтийг зайлсхийхийн тулд мөрдөх чадвартай холбоотой өгөгдлийг зохих ёсоор хадгалж, салбарт хамааралтай хугацаанд хадгалах.
- 3.38 Мөрдөх чадварын арга хэмжээнүүд нь их хэмжээний үйл ажиллагааны өгөгдлийг хуримтлуулж болзошгүй учраас байгууллагууд бүтээгдэхүүний аль шинж чанарууд нь мөрдөх чадвар шаардлагатай байгааг, мөн аль мөрдөх чадварын арга хэмжээнүүд тэдний хэрэгцээнд хангалттай болохыг, хиймэл оюун ухааны загварын шийдвэрүүд, өгөгдлийн багцууд болон процессуудыг баримтжуулахад шаардлагатай нөөцийг харгалзан үзэж болно. Байгууллагууд үүнийг дараах хэд хэдэн хүчин зүйлсийн үндсэн дээр үнэлж болно:
- a. Хиймэл оюун ухааны системийг ашигласнаас үүсэх хохирлын магадлал болон/эсвэл ноцтой байдлын үнэлгээ;
  - b. Хиймэл оюун ухааны загварыг өмнө нь хэр их туршиж үзсэн эсвэл ашигласан байдал; болон
  - c. Тэдний салбарын зохицуулалтын хэрэгцээ.

<sup>9</sup> Анхаарах ёстой зүйл бол "хар хайрцаг" бичигч нь хиймэл оюун ухааны загварын утгаар "хар хайрцаг" биш юм (өөрөөр хэлбэл, хиймэл оюун ухааны загварын шийдвэр гаргах үйл явц нь үндсэндээ тайлбарлахад хэцүү, ойлгомжгүй байдаг).

### Давтагдах чадвар

3.39 Давтагдах чадвар нь тухайн байгууллагын доторх үр дүнг давтах чадварыг хэлдэг бол нөхөн үйлдвэрлэх чадвар нь бие даасан баталгаажуулалтын баг ижил хиймэл оюун ухааны аргыг ашиглан байгууллагын баримтжуулсан мэдээлэлд үндэслэн ижил үр дүн гаргах чадварыг хэлнэ. Нөхөн үйлдвэрлэх чадвар нь хиймэл оюун ухааны бүтээгдэхүүний болон түүнийг хэрэгжүүлж буй байгууллагын найдвартай байдалд нөлөөлж чадна. Нөхөн үйлдвэрлэх чадварыг хэрэгжүүлэх нь гадаад талуудын оролцоог шаарддаг тул байгууллагууд бүтээгдэхүүн, үйлчилгээндээ гадны нөхөн үйлдвэрлэх чадварын тест шаардлагатай хиймэл оюун ухааны онцлог шинж чанаруудыг тодорхойлохдоо эрсдэлд суурилсан хандлагыг авч үзэж болно.

3.40 Давтагдах чадварыг дэмжих дараах практикууд:

- a. Давтагдах чадварыг хангахын тулд тодорхой нөхцөл байдал эсвэл тусгай нөхцлүүдийг шалгаж үзэх;
- b. Хиймэл оюун ухааны загварын найдвартай байдал болон давтагдах чадварыг баталгаажуулах баталгаажуулалтын аргуудыг хэрэгжүүлэх;
- c. Шалгалт болон нөхөн үйлдвэрлэх үйл явцыг хөнгөвчлөхийн тулд хиймэл оюун ухааны загварын хөгжүүлэлтийн процессын алхам бүрийг давтаж үзэх файлуудыг бэлэн байлгах;
- d. Арилжааны зориулалттай, бэлэн хиймэл оюун ухааны системийг худалдан авдаг компаниуд анхны хиймэл оюун ухааны шийдэл нийлүүлэгчээс загварын үр дүн нөхөн үйлдвэрлэгдэж байгаа эсэхийг шалгах;
- e. 3.30 (c)-(e) догол мөрөнд дурдсан давтагдах чадварын заалтуудыг хэрэгжүүлэх (тухайлбал, онцгой тохиолдлуудыг хэрхэн илрүүлэх болон зохицуулах, онцгой тохиолдлуудыг зохицуулах нь байгууллагын бодлоготой нийцэж байгаа эсэхийг баталгаажуулах, мөн цаг хугацааны явцад гарч буй өөрчлөлтийг тодорхойлж, тооцох).

## Аудит хийх чадвар

- 3.41 Аудит хийх чадвар гэдэг нь хиймэл оюун ухааны системийн алгоритмууд, өгөгдөл болон загварын процессуудыг үнэлэхэд бэлэн байдлыг хэлнэ. Дотоод болон гадаад аудиторуудын үнэлгээ (мөн үнэлгээний тайлангийн боломжтой байдал) нь хиймэл оюун ухааны системийн найдвартай байдлыг нэмэгдүүлж, загварын хариуцлага, практик, үр дүнгийн зөвтгөлтэй байдал зэргийг харуулдаг. Гэхдээ, аудит хийх чадвар нь бизнесийн загварууд эсвэл хиймэл оюун ухаантай холбоотой оюуны өмчийн мэдээллийг нийтэд ил болгох шаардлагагүй.
- 3.42 Аудит хийх чадварыг хэрэгжүүлэх нь зөвхөн гадаад талуудын оролцоог шаарддаггүй, мөн аудиторууд (гаднынх байж болох) арилжааны мэдрэмтгий мэдээллийг задруулах шаардлагатай болдог. Байгууллагууд бүтээгдэхүүн, үйлчилгээндээ аудит хийх чадвар хэрэгжүүлэх шаардлагатай хиймэл оюун ухааны онцлог шинж чанаруудыг тодорхойлохдоо эрсдэлд суурилсан хандлагыг авч үзэж болно, эсвэл зохицуулалтын шаардлагуудтай нийцүүлэх шаардлагатай үед хэрэгжүүлэх шаардлагатай.
- 3.43 Аудит хийх чадварыг хөнгөвчлөхийн тулд байгууллагууд өгөгдлийн гарал үүсэл, худалдан авалт, урьдчилсан боловсруулалт, угсаа гаргалгаа, хадгалалт, аюулгүй байдлын бүрэн бүртгэлийг хадгалах боломжтой. Энэхүү бүртгэлд өгөгдлийн төлөөлөл, өгөгдлийн хангалттай байдал, эх үүсвэрийн бүрэн бүтэн байдал, өгөгдлийн цаг хугацааны мөр, өгөгдлийн хамаарал, ажлын урсгалд гарсан урьдчилан хараагүй өгөгдлийн асуудлууд зэрэг чанарын оролтыг багтааж болно.
- 3.44 Байгууллагууд ийм мэдээллийг төвлөрүүлэн дижитал процессын бүртгэлд хадгалахыг хүсэж болно. Энэ нь байгууллагад нэг дор мэдээллийг хялбар байлгах, холбогдох талуудад болон шийдвэр гаргагчдад загварын хариуцлага, практик, системийн үйл ажиллагааны үр дүнгийн зөвтгөлтэй байдлыг харуулахад туслах боломжийг олгодог. Ийм бүртгэл нь мэдээллийн хандалтыг зохион байгуулах, олон нийтээс хамгаалах ёстой өгөгдлийг хадгалах, үр дүнгийн танилцуулгыг өөр өөр сонирхол, мэргэжлийн түвшний оролцогчдод тохируулах боломжийг нэмэгдүүлдэг.

# SYMPHONY AYASDI AI:

## Загвар Хөгжүүлэлтийн Баримтжуулалтын Талаарх Зураглал

Symphony Ayasdi AI ("Ayasdi") нь АНУ-ын банк санхүүгийн салбар дахь үйлчлүүлэгчдэдээ эдийн засгийн стресс нөхцөлд алдагдлыг шингээхэд шаардлагатай хөрөнгийн нөөцийг таамаглах чадвартай хиймэл оюун ухааны загваруудыг бүтээхэд тусалдаг шийдлийг санал болгодог. Үйлчлүүлэгчид нь АНУ-ын Холбооны Нөөцийн Системд (Federal Reserve) өөрсдийн хиймэл оюун ухааны загваруудын үнэн зөв болон хамгаалагдах чадварыг нотлох шаардлагатай байдаг.

Шийдэл нь Ayasdi-ийн Model Accelerator (АМА) юм. Энэ нь эхлээд загварт оруулах шаардлагатай хувьсагчдыг тодорхойлж, дараа нь яагаад тэдгээрийг сонгосныг тайлбарлана. АМА нь үүнийг өргөтгөсөн анхан шатны түвшний өгөгдөлд кодлогдсон боломжит харилцааг судалж, нуугдсан, урьдчилан таамаглах үнэ цэнэтэй хэв маягийг олж тодорхойлсноор гүйцэтгэдэг. АМА нь сонгосон хувьсагчдыг ашиглан хиймэл оюун ухааны загваруудыг бүтээдэг. Энэ нь хэрэглэгчдийн сонголтод зориулан нэг үндсэн загвар болон хэд хэдэн боломжит өрсөлдөх загварыг санал болгодог. Хэрэглэгчдийн байгууллагын бизнес нэгжүүд үндсэн болон өрсөлдөх загваруудыг үнэлж, тэдний бизнес нэгжүүдэд хамгийн сайн тохирох загваруудыг сонгох болно.

Загвар бүтээх бүх үйл явц автоматаар баримтжуулдаг. Хэрэглэгчид АМА-ыг ашиглан хувьсагчдыг сонгох, загвар бүтээх арга зүйгээ системчилж, тасралтгүй тайлангаар загварын угшил, хувьсагчдыг сонгох, кросс баталгаажуулалтын талаар дэмжлэг үзүүлэх боломжтой. Энэ нь хэрэглэгчдэд анхны онцлог, загварын сонголтуудыг бүртгэж, баримтжуулах боломжийг олгодог. Үүний зэрэгцээ, загвар бүтээх, батлах бүх үйл явцыг хянаж, каталогийн аргаар бүртгэж, дахин ашиглахад хялбар болгодог.

Загвар бүтээх нарийвчилсан үйл явцыг үзүүлэх, өрсөлдөх загваруудыг үнэлэх хатуу чанга арга зүй нь Ayasdi-ийн хэрэглэгчдэд АНУ-ын Холбооны Нөөцийн Газарт эцсийн загваруудыг хэрхэн сонгосноо тайлбарлах боломжийг олгодог.



## ОРОЛЦОГЧДЫН ХАРИЛЦАА БА ХАРИЛЦАА ХОЛБОО

3.45 Энэ хэсэг нь байгууллагуудад хиймэл оюун ухааныг нэвтрүүлэх үед оролцогч талуудын харилцааны стратегид итгэлцлийг бий болгоход зохих алхмуудыг авахад туслах зорилготой юм.

### Ерөнхий мэдээлэл

3.46 Байгууллагууд бүтээгдэхүүн болон/эсвэл үйлчилгээндээ хиймэл оюун ухааныг ашиглаж байгаа эсэх талаараа ерөнхий мэдээллийг өгөхийг уриалж байна. Боломжтой бол энэ нь хиймэл оюун ухаан гэж юу болох, хэрэглэгчидтэй холбоотой шийдвэр гаргалтад хэрхэн ашиглагддаг, ямар ашиг тустай болох, байгууллага нь яагаад хиймэл оюун ухааныг ашиглах шийдвэр гаргасан, эрсдлийг бууруулахын тулд ямар алхмуудыг хийсэн, хиймэл оюун ухаан нь шийдвэр гаргалтын үйл явцад ямар үүрэг, хэмжээтэй оролцдог талаар мэдээлэл агуулж болно. Жишээлбэл, онлайн портал нь хэрэглэгчдэдээ хиймэл оюун ухааныг ашигласан чатботтой харилцаж байгаа бөгөөд хүний үйлчилгээний ажилтантай биш гэдгийг мэдээлж болно.

3.47 Байгууллагууд хиймэл оюун ухааны шийдвэр нь хувь хэрэглэгчдэд хэрхэн нөлөөлж болох, мөн шийдвэрийг буцааж болох эсэх талаар мэдээлэл өгөхийг бодож үзэж болно. Жишээлбэл, байгууллага нь хэрэглэгчдэд зээлийн үнэлгээ нь зөвхөн энэ байгууллагаас зээл авах татгалзалд хүргэхээс гадна ижил төстэй бусад байгууллагуудаас зээл авах татгалзалд хүргэж болохыг, мөн хэрэв хэрэглэгчид зээлийн чадварын талаар илүү нотолгоо гаргаж чадвал ийм шийдвэрийг буцааж болохыг мэдээлж болно.

### Тайлбар өгөх бодлого

3.48 Байгууллагууд хувь хүмүүст ямар тайлбар өгөх, тэдгээрийг хэзээ өгөх талаар бодлого боловсруулахыг уриалж байна. Ийм бодлого нь харилцаанд нийцтэй байдлыг хангаж, байгууллагын янз бүрийн гишүүдийн үүрэг, хариуцлагыг тодорхой зааж өгдөг. Үүнд хиймэл оюун ухаанаар нэмэгдсэн шийдвэр гаргалтын үйл явц хэрхэн ажилладаг, тодорхой шийдвэрийг хэрхэн гаргасан болон уг шийдвэрийн ард байгаа шалтгаан, уг шийдвэрийн нөлөө ба үр дагаврын талаар тайлбар өгөхийг багтааж болно. Тайлбар нь ерөнхий харилцааны нэг хэсэг байж болно. Энэ нь мөн хүсэлтээр тодорхой шийдвэрийн талаарх мэдээлэл байж болно. Энэ талаар, тэнцүү байдлын зарчим нь хүний удирдлагаар гарсан шийдвэрүүдэд хамаарах ил тод байдлын нэгэн адил хиймэл оюун ухааны системээр гарсан эсвэл нэмэгдсэн шийдвэрүүдэд мөн хамаардаг.



### Тайлбар өгөх болон ил тод байдлыг утга учиртайгаар нэгтгэх нь

- 3.49 Зохистой харилцаа ба харилцаа холбоо нь байгууллагууд болон хувь хүмүүс (үүнд ажилчид орно) хооронд нээлттэй харилцааг бий болгож, хадгалснаар итгэл ба итгэлцлийг төрүүлдэг. Оролцогч талуудын харилцааны стратеги нь нэг хэвээр үлдэх ёсгүй. Компаниуд стратегиудынхаа үр нөлөөг шалгах, үнэлэх, хянахыг уриалж байна. Цаашилбал, эдгээр хүчин зүйлсийг хэрэгжүүлэх хэмжээ, арга хэлбэр нь нөхцөл байдлаас хамааран өөр өөр байж болно.
- 3.50 Олон төрлийн оролцогч талуудын мэдээллийн хэрэгцээ өөр өөр байдаг тул байгууллага нь эхлээд өөрийн үзэгчдийг (өөрөөр хэлбэл, гадаад болон дотоод оролцогч талууд) тодорхойлох боломжтой. Байгууллагын гадаад оролцогч талуудад хэрэглэгчид, зохицуулагчид, хамтран ажиллаж буй бусад байгууллагууд, нийт нийгэм орно. Харин дотоод оролцогч талуудад байгууллагын удирдах зөвлөл, удирдлага болон ажилчид орно. Байгууллага мөн оролцогч талуудтай харилцах зорилго болон нөхцөлийг авч үзэж болно. Жишээ татах үүднээс, энэхүү Загвар Хүрээ нь хэрэглэгчид болон бусад байгууллагуудтай харилцах үед авч үзэх асуудлуудыг танилцуулж байна.

*Оролцогч талуудын мэдээллийн хэрэгцээ өөр өөр байдаг тул байгууллага нь эхлээд өөрийн үзэгчдийг тодорхойлж, оролцогч талуудтай харилцах зорилго болон нөхцөлийг авч үзэж болно.*

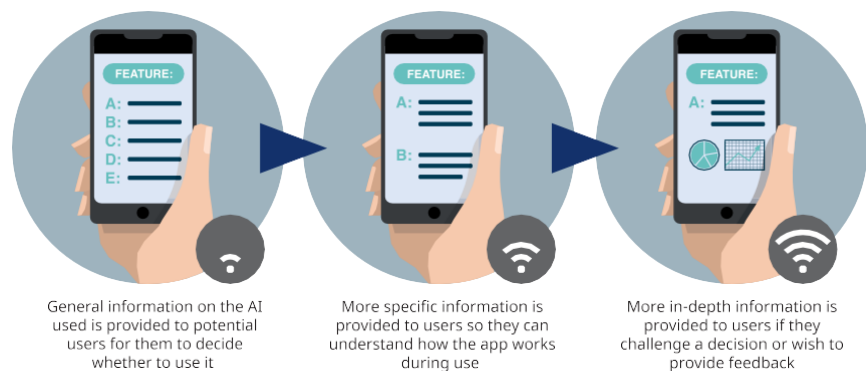
### Хэрэглэгчидтэй харилцах нь

- 3.51 Байгууллагуудыг хэрэглэгчдийн хиймэл оюун ухаантай харилцах аяллын явцад тэдний мэдээллийн хэрэгцээг авч үзэхийг уриалж байна. Энэ нь хиймэл оюун ухааны шийдлийг ашиглах эсэхээ шийдэхээс эхлээд, ашиглах үедээ хиймэл оюун ухааны шийдэл хэрхэн ажиллаж байгааг ойлгох, мөн хиймэл оюун ухааны шийдвэрийг дахин шалгуулах хүсэлт гаргах хүртэлх алхмуудыг хамардаг. Хэрэглэгчдийн ердийн аялал нь дараах мэдээллийн хэрэгцээг хангахыг шаардаж болно:

- a. Хэрэглэгчид сонгож буй бүтээгдэхүүн эсвэл үйлчилгээ нь хиймэл оюун ухаан ашигладаг болохыг мэдэж байгаа эсэхийг баталгаажуулах. Ийм мэдээллийг ерөнхий бүтээгдэхүүний тодорхойлолтод багтаан өгөх боломжтой.
- b. Хэрэглэгчид хиймэл оюун ухаан ашигласан онцлог шинжүүдийг энгийн хэрэглээний үед хэрхэн ажиллахыг мэдэх боломжтой байх үүднээс мэдээлэл өгөх. Энэхүү мэдээллийг бүтээгдэхүүний шинж чанаруудын дэлгэрэнгүй тайлбар эсвэл тодорхойлолтоор дамжуулан өгөх боломжтой.

Гэхдээ энэ нь хиймэл оюун ухаан ашигласан бүх онцлог шинжүүдэд заавал шаардлагатай биш байж болно. Байгууллагууд хэрэглэгчийн итгэлийг нэмэгдүүлэхэд нэмэлт мэдээлэл өгөх нь ач холбогдолтой онцлог шинжүүдийг тодорхойлохыг уриалж байна. Түүнчлэн, хэрэв хиймэл оюун ухааныг шийдвэр гаргахад ашиглаж байгаа бол хэрэглэгчид хиймэл оюун ухааны тусламжтайгаар гаргасан шийдвэрүүд тэдэнд хэрхэн нөлөөлж болохыг ойлгох үүднээс мэдээлэл өгөх боломжтой. Үүнийг үйлчилгээ хэрхэн үзүүлэх талаар тайлбарлах замаар мөн өгөх боломжтой.

- c. Хэрэглэгчидтэй байнга харилцдаг хиймэл оюун ухаан ашигласан онцлог шинжүүдийн хувьд, хэрэглэгчид яагаад хиймэл оюун ухаан ашигласан онцлог нь тодорхой байдлаар ажиллаж байгааг ойлгох боломжтой мэдээлэл өгөх, мөн боломжтой бол хэрэглэгчид ирээдүйн үйл ажиллагаанд нөлөөлөх боломжтой тохиргоонуудыг санал болгох. Үүнийг хийх нь илүү их инженерийн хүчин чармайлт шаарддаг (жишээлбэл, хэрэглэгчийн түүхийг харуулсан нэмэлт хэрэглэгчийн интерфэйсийг бий болгох гэх мэт) бөгөөд өгөгдөх мэдээллийн түвшин нь онцлог шинжүүдийн тайлбаруудаас илүү дэлгэрэнгүй, хувьчилсан байж болох тул байгууллагууд бүтээгдэхүүнийхээ аль онцлог шинжүүдийг энэ түвшний дэлгэрэнгүй мэдээллээр хангахад ашиг тустайг шийдэх хэрэгтэй болно.
- d. Хэрэглэгчдэд нөлөөлөх хиймэл оюун ухаанаар нэмэгдсэн шийдвэрүүдийн хувьд, шийдвэрүүд яагаад гарсан талаар ойлгоход туслах нэмэлт мэдээлэл өгөхийг бодож үзэх; мөн ийм шийдвэрүүдийн зарим ангиллын хувьд, шийдвэрийг маргах боломжтой зохистой сувгийг санал болгох. Өгөгдөх мэдээллийн түвшин дэлгэрэнгүй байх шаардлагатай боловч хэрэглэгч шийдвэрээс нөлөөлсөн тохиолдлуудад л шаардлагатай байж болох юм.



### Сонголтыг татгалзах боломж

- 3.52 Байгууллагууд хиймэл оюун ухаан ашигласан бүтээгдэхүүн эсвэл үйлчилгээнээс татгалзах сонголтыг хувь хүмүүст өгөх эсэх, мөн энэ сонголтыг анхдагчаар санал болгох уу эсвэл зөвхөн хүсэлтийн дагуу санал болгох уу гэдгийг шийдэхдээ сайтар анхаарах хэрэгтэй. Холбогдох хүчин зүйлүүдэд дараах зүйлс орно:
- Хувь хүмүүст учирч болох эрсдэл/хор хөнөөлийн түвшин;
  - Гаргасан шийдвэрийн буцаах боломж;
  - Өөр шийдвэр гаргах механизмын боломж;
  - Өөр механизмын зардал эсвэл үр дагавар;
  - Хоёрдогч системүүдийг хадгалах төвөгтэй байдал, үр ашиггүй байдал;
  - Техникийн боломжит байдал.
- 3.53 Байгууллага дээрх хүчин зүйлсийг үнэлж, татгалзах сонголтыг өгөхгүй гэж шийдсэн тохиолдолд, хэрэглэгчдэд шийдвэрийг дахин хянах сувгийг санал болгох зэрэг арга хэмжээг авч үзэх нь зохистой байдаг. Шаардлагатай тохиолдолд, байгууллагууд гомдол хүлээн авах эсвэл хэрэглэгчдээс нөхөн олговор авах үед чатботын ярианы түүхийг хадгалахыг хүсэж болно.

### Харилцаа холбооны сувгууд

- 3.54 Байгууллагууд хэрэглэгчдэдээ зориулан дараах харилцаа холбооны сувгуудыг хэрэгжүүлэхийг уриалж байна:

### Feedback channels

a.

Энэхүү суваг нь хэрэглэгчдэд санал хүсэлт гаргах эсвэл асуулт тавих боломжийг олгоно. Хэрэв тохиромжтой бол энэ сувгийг байгууллагын Мэдээллийн хамгаалалтын ажилтан (DPO) удирдаж болно. Хэрэглэгчид өөрсдөд нь нөлөөлсөн шийдвэрүүдэд ашиглагдсан хувийн мэдээлэлд алдаа илрүүлсэн тохиолдолд, энэ суваг нь тэдний мэдээллийг засах боломжийг олгоно. Ийм засвар болон санал хүсэлт нь мэдээллийн үнэн зөв байдлыг хадгалах болно. Хэрэв оролцогч талууд өөрсдийн тухай гаргасан материаллаг дүгнэлтэд санал хүсэлт болон асуулт тавихыг хүсвэл, энэ сувгийг байгууллагын Чанарын үйлчилгээний менежер (QSM) удирдаж болно.

### Decision review channels

b.

Одоогийн хяналтын үүрэг хариуцлагаас гадна байгууллагууд хувь хүмүүс (жишээ нь, хохирол амссан хэрэглэгч) өөрт нь нөлөөлсөн хиймэл оюун ухааны чухал шийдвэрийг дахин хянах хүсэлт гаргах боломжийг олгохыг бодож үзэж болно. Хэрэв бүрэн автомат шийдвэр нь хэрэглэгчид материаллаг нөлөө үзүүлж болзошгүй бол уг шийдвэрийг хүнээр дахин хянах боломжийг олгох нь зүйтэй юм.

## Хэрэглэгчийн интерфэйсийг шалгах

3.55 Байгууллагуудыг хэрэглэгчийн интерфэйсийг нэвтрүүлэхээс өмнө шалгаж, ашиглахад тохиромжгүй асуудлуудыг шийдвэрлэхийг уриалж байна. Ингэснээр хэрэглэгчийн интерфэйс нь өөрийн зорилгодоо нийцсэнээр үйлчилнэ. Хэрэв боломжтой бол байгууллагууд хувь хүмүүст тэдний хариултыг хиймэл оюун ухааны системийг сургахад ашиглах болно гэдгийг мэдээлэхийг уриалж байна (жишээ нь, чатбот). Байгууллагууд ийм хариултыг ашиглах эрсдэлээс болгоомжлох хэрэгтэй, учир нь зарим хүмүүс зориудаар “муу үг хэллэг” эсвэл “санамсаргүй хариулт” ашиглаж, хиймэл оюун ухааны системийн сургалтад нөлөөлж болзошгүй юм.

## Ойлгомжтой харилцаа холбоо

3.56 Байгууллагууд ил тод байдлыг нэмэгдүүлэхийн тулд ойлгомжтой байдлаар харилцахыг уриалж байна. Одоогийн байдлаар ойлгомжтой байдлыг хэмжих хэрэгслүүд байдаг бөгөөд үүнд ойлгомжтой байдлын график, Gunning Fog индекс, Flesch-Kincaid ойлгомжтой байдлын тест гэх мэт орно. Өндөр нөлөөтэй шийдвэрүүдийг ойлгомжтой байдлаар харилцаж, ашиглаж буй технологийн талаар ил тод байх нь тус болно. Текстэн харилцаанаас гадна байгууллагууд дүрслэлийн хэрэгслүүд, график төлөөлөл, хураангуй хүснэгтүүд эсвэл эдгээрийн хослолыг ашиглахыг бодож үзэж болно. Зорилго нь тайлбар эсвэл тайлбарыг байгууллагын хэрэглэгчид болон бусад оролцогч талуудад ойлгомжтой байдлаар дамжуулах явдал юм.

### Хэрэглэгчийн хүлээн зөвшөөрөгдсөн бодлого

- 3.57 Зарим тохиолдолд байгууллагууд бодит амьдралын өгөгдлөөр сургадаг хиймэл оюун ухааны шийдлүүдийг (жишээ нь, идэвхтэй сургалт) хэрэгжүүлж байж болно. Эдгээр байгууллагууд хэрэглэгчид шийдлийн загварын гүйцэтгэл, үр дүнд хортойгоор нөлөөлж болзошгүй оролтын өгөгдлийг санаатайгаар оруулахгүй байхыг баталгаажуулах зорилгоор тодорхой хүлээн зөвшөөрөгдсөн хэрэглэгчийн бодлого (“AUPs”) тогтоох талаар бодож үзэж болно. Энэ нь өнгөрсөн үеийн AI чатбот системүүдийн жишээгээр үндэслэлгүй хортойгоор зүй бусаар удирдуулж, олон нийтэд хүлээн зөвшөөрөгдөөгүй хариу өгөх болсон тохиолдлуудаас үүдэн гарч байна.
- 3.58 Энэ талаар, AUPs нь хувь хүмүүс AI системтэй хийх харилцан үйлчлэлийн өргөн хязгаарыг тогтоох үүрэгтэй бөгөөд энэ нь зориудаар үйлдэл хийх, эсвэл урвуу инженерчлэл хийх, идэвхгүй болгох, хөндлөнгөөс оролцох, хиймэл оюун ухаан ашигласан үйлчилгээний үйл ажиллагаа, бүрэн бүтэн байдал эсвэл гүйцэтгэлд саад учруулах оролдлогуудтай холбоотой хязгаарлалтуудыг багтаадаг.

### Бусад байгууллагуудтай харилцах нь

- 3.59 Өмнөх хэсэгт тайлбарласан зарим арга барил, арга зүй нь байгууллагууд хиймэл оюун ухааны шийдэл нийлүүлэгчидтэй (жишээ нь, хиймэл оюун ухааны шийдлийг худалдан авах, зохицуулалтын зөвшөөрөл авах), эсвэл бусад байгууллагуудтай (жишээлбэл, салбарын хамтын ажиллагааг хөнгөвчлөх, системүүдийн уялдаа холбоог хангах) харилцах үед мөн хамаатай байдаг. Тиймээс байгууллагууд бизнесийн зорилгодоо хүрэхэд туслахын тулд хиймэл оюун ухааны шийдэл нийлүүлэгчдээс хангалттай мэдээлэл авах шаардлагатай (жишээлбэл, 3.51-р догол мөрөнд тайлбарласан мэдээллийг хангах талаар урт хугацааны тохиролцоо). Энэ нь хиймэл оюун ухааны шийдэл нийлүүлэгчдээс мэдээлэл өгөх<sup>10</sup> болон байршуулж буй байгууллага нь Загвар Хүрээтэй нийцэх боломжийг бүрдүүлэх шаардлагатай онцлогийг<sup>11</sup> бий болгоход дэмжлэг авах зэрэг энгийн байх боломжтой.

<sup>10</sup> Жишээлбэл, сургалтын өгөгдлийн багцын угшлын талаархи мэдээлэл, загварын сургалт болон сонголтын үйл явцын гол алхмуудыг баримтжуулах зэрэг мэдээлэл.

<sup>11</sup> Жишээлбэл, хиймэл оюун ухаан ашигласан онцлог шинжийн хүлээгдэж буй үйл ажиллагааны талаар мэдээлэл өгч, хэрэглэгчид ирээдүйд хиймэл оюун ухаан ашигласан онцлог шинжийн гүйцэтгэлд нөлөөлөх тохиргоог удирдах боломжийг олгох функц бүтээх зэрэг хэрэглэгчидтэй харилцах харилцаанууд.

- 3.60 Байгууллагууд AI шийдэл нийлүүлэгчдээс дараах зүйлсийн талаар дэмжлэг болон дэлгэрэнгүй мэдээллийг авах шаардлагатай байж болно:
- a. Өгөгдөл (жишээ нь, алгоритмыг сургахад ашигласан өгөгдлийн төрлүүд ба хүрээ, гадаад сургалтын өгөгдлийн эх үүсвэр ба чанар);
  - b. Загварын сургалт ба сонголт (жишээ нь, ашигласан онцлог шинжүүд ба хувьсагчид, нийлүүлсэн арилжааны загваруудын жингүүд, загварын сургалт болон сонголттой холбоотой гаргасан гол шийдвэрүүдийг баримтжуулах);
  - c. Хүний оролцоо (жишээ нь, алгоритм боловсруулах болон шийдвэр гаргах үйл явцад хүний оролцооны шинж чанар);
  - d. Дүгнэлтүүд (жишээ нь, алгоритмын гаргасан таамаглалууд ба эдгээрийг бүтээгдэхүүний онцлог шинжүүд эсвэл шийдвэр гаргалтад хэрхэн тусгасан);
  - e. Алгоритмын оролцоо (жишээ нь, шийдлийн аль хэсэгт алгоритмыг ашигласан);
  - f. Өгөгдөл ба алгоритмын алдааг багасгах арга хэмжээ болон хамгаалалт.
- 3.61 Зорилго болон нөхцөл байдлаас хамааран шаардлагатай мэдээллийн төрөл болон дэлгэрэнгүй түвшин өөр байж болно. Жишээлбэл, зохицуулагч нь зохицуулалттай байгууллагаас загвар боловсруулах болон сонгох үйл явц хангалттай нарийвчилсан болохыг нотлохыг шаарддаг байж магадгүй бөгөөд AI шийдэл нийлүүлэгчээс илүү их мэдээлэл өгөхийг шаарддаг ба зохицуулагчтай тодруулга хийх үйл явцад оролцох шаардлагатай байж болно. Харин салбарын хамтрагч нь нийцэл болон уялдаа холбоотой холбоотой хүчин зүйлст илүү анхаарал хандуулж байж магадгүй.

### Ёс зүйн үнэлгээ

- 3.62 Эцэст нь, хиймэл оюун ухааны хөгжүүлэлт ба хэрэглээг зохицуулдаг ёс зүйн стандартууд хөгжихийн хэрээр байгууллагууд өөрсдийн хиймэл оюун ухааны удирдлагын практик болон үйл явц нь хөгжиж буй AI стандартуудтай нийцэж байгаа эсэхийг үнэлэхийг уриалж байна, мөн эдгээр үнэлгээний үр дүнг холбогдох оролцогч талуудад нээлттэй байлгахыг зөвлөж байна.

# FACEBOOK:

## ОРОЛЦОГЧ ТАЛУУДЫН ХАРИЛЦАА БА ХАРИЛЦААНЫ ЖИШЭЭ

Нийгмийн сүлжээ ба технологийн компани болох Facebook нь өөрийн үйл ажиллагаа болон үйлчилгээндээ хиймэл оюун ухааныг ашиглах талаар олон нийт болон хэрэглэгчдэд ил тод байхыг зорьж байна.

Facebook нь хэрэглэгчидтэй ил тод байхыг зорьдог ба үүнийг дараах байдлаар хэрэгжүүлдэг:

- a. Facebook-ийн өгөгдөл цуглуулалт, ашиглалтыг ойлгомжтой байдлаар нийтэд мэдээлэх. Энэ нь Үйлчилгээний Нөхцөл болон Өгөгдлийн Бодлогын дагуу, тайлбар бүхий хэрэглэгчдэд ээлтэй видео бичлэгүүдээр дамжуулан хийгддэг;
- b. Хэрэглэгчдэд өөрсдийн мэдээллийг хэрхэн ашиглаж, хуваалцах талаар хялбар бөгөөд утга учиртай хяналт өгөх;
- c. Нийтлэл цувралаар дамжуулан бодлогыг тайлбарлах, нарийн төвөгтэй сэдвүүдийг хэлэлцэх, Facebook-ийн шийдвэрүүдийн үндэслэлийг тайлбарлах, мэргэжилтнүүдийг санал бодлоо хуваалцахыг урьсан блог бичлэгүүдийг нийтлэх. Жишээлбэл, Facebook нь "Hard Questions" нэртэй блог бичлэгээр Нүүр Царай Таних технологийн талаар хэлэлцсэн бөгөөд хэрхэн хэрэглэгчдийг зураг тэмдэглэхэд тусалдаг, мөн хэрэгжүүлсэн хяналтуудыг тайлбарласан;
- d. Хэрэглэгчдэд янз бүрийн бүтээгдэхүүн, онцлог шинжүүдийн үндсэн технологийг ойлгуулах зорилготой хиймэл оюун ухааны боловсролын санаачилга, кампанит ажлуудыг сурталчлах. Жишээлбэл, Facebook-ийн Хиймэл Оюун Ухааны Судалгааны Лаборатори нь Машин Сургалтын алгоритмууд болон тэдгээрийг хэрхэн ашиглаж байгааг тайлбарлах хиймэл оюун ухааны боловсролын видео бичлэгүүдийг боловсруулж, нийтэлсэн.

Facebook нь одоогоор хэрэглэгчдэд хамгийн тохиромжтой постуудыг харуулдаг өөрчилсөн News Feed үйлчилгээг үзүүлдэг. News Feed-ын агуулгыг хэрэглэгч найзлах, дагахыг сонгосон хүмүүс болон хуудсууд тодорхойлдог.

News Feed-н онцлогийг зөвлөх ажлын хүрээнд Facebook дараах зүйлсийг авч үзсэн:

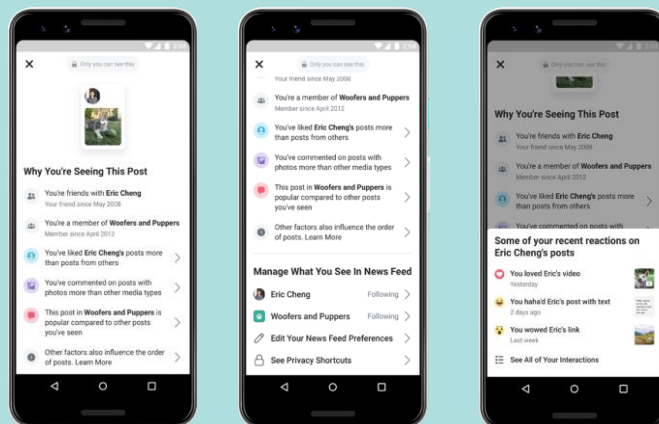
- a. Онцлогийн ард буй алгоритмуудын талаар ил тод байх, илүү их мэдээлэл өгөх шаардлага;



- b. Хэрэглэгчдэд хамгийн их үнэ цэнэтэй мэдээллийн төрлүүд. Жишээлбэл, Facebook нь News Feed-д оруулсан постуудын ард буй хүмүүсийн харилцан үйлчлэлийн жишээнүүдийг оруулсан;
- c. Хэрэглэгчдэд News Feed-ээ удирдах, хянах шаардлага.

Үүний үндсэн дээр, Facebook хэрэглэгчдийн итгэлийг бий болгохын тулд дараах арга хэмжээг авч хэрэгжүүлсэн:

- a. “Яагаад энэ постыг харж байна вэ?” онцлогийг хэрэгжүүлж, хэрэглэгчдийн өнгөрсөн харилцаа News Feed дахь постуудын эрэмбэлэлтэд хэрхэн нөлөөлсөн талаар тайлбарлах. Тодруулбал, хэрэглэгчид дараах зүйлсийг сурах боломжтой болсон:
  - i. News Feed дахь тодорхой постыг харах шалтгаан. Жишээлбэл, пост нь хэрэглэгчийн найз эсвэл хэрэглэгчийн дагасан Групп эсвэл Хуудаснаас байж болно.
  - ii. Постуудын дараалалд хамгийн их нөлөөлсөн мэдээлэл, үүнд: (а) хэрэглэгч хүмүүс, Хуудас эсвэл Группын постуудтай хэр давтамжтай харилцаж байсан; (б) хэрэглэгч тодорхой төрлийн постуудтай (жишээлбэл, видео, зураг эсвэл холбоосууд) хэр давтамжтай харилцаж байсан; (в) хэрэглэгчийн дагасан хүмүүс, Хуудас болон Группээс хуваалцсан постуудын алдартай байдал.
  - iii. News Feed-ээ хувийн болгож, See First, Unfollow, News Feed Preferences болон Privacy Shortcuts зэрэг хяналтын товчлолууд.
- b. News Feed-н гол шинэчлэлтүүдийн үндэслэлийг онцлон тайлбарласан “News Feed FYI” блог бичлэг цувралыг нийтлэх.
- c. Facebook-ийн системүүд хэрхэн ажилладаг болон Facebook-ийн өөрчлөлтүүдийг хэрхэн үнэлдэг талаар илүү дэлгэрэнгүй мэдээлэл өгсөн шинэ “Inside Feed” вэбсайтыг нээх.



# MSD:

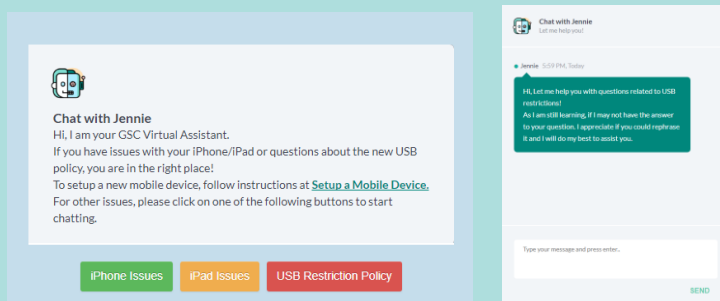
## ОРОЛЦОГЧ ТАЛУУДТАЙ ХАРИЛЦАХ БОЛОН ХАРИЛЦААНЫ ЖИШЭЭ

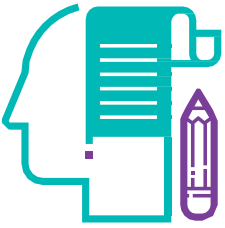
MSD бол мэдээлэл технологийн асуудлаар асуултад хариулах зориулалттай дотоодын чатбот болох Jennie-г ашигладаг олон улсын эмийн компани юм.

Нэвтрүүлэхээс өмнө, MSD-ийн Хэрэглэгчийн Туршлага (UX) баг хүний-Хиймэл Оюун Ухааны интерфэйсийг шалгаж, хэрэглэгчийн харилцан үйлчлэлтийг сайжруулах зорилгоор ашиглахад тохиромжгүй асуудлуудыг шийдвэрлэсэн. Тодруулбал, Jennie-ийн хөгжүүлэлт болон нэвтрүүлэлтийг дараах гурван зарчим удирдсан:

- Хэрэглэгчийн сэтгэхүйн загварыг ойлгох:** UX баг төлөөлөх хэрэглэгчидтэй хэрэглэгчийн судалгаа явуулж, чатботтой харилцах үед хэрэглэгчдийн хүлээлтийг ойлгохыг зорьсон. Судалгаанд мэдээллийн технологийн асуултуудын хүрээ, хүлээгдэж буй хариултууд болон өгөгдөх хариултуудын төрлүүд (жишээ нь, хариултууд хэрхэн техник хэлийг ашиглах) багтсан. Хэрэглэгчдэд ээлтэй интерфэйс бүтээснээр, хэрэглэгчид чатботтой харилцахад илүү тав тухтай байх болно.
- Хүн төвтэй хандлага баримтлах:** Хүний зан төлөвийн хэв маягийг ойлгохын тулд баг ажилчид нь чатботтой харилцахдаа хэрхэн хариу үйлдэл үзүүлж байгааг шинжилсэн. Үүнд хэрэглэгчид асуултуудаа хэрхэн томъёолдог, ямар төрлийн хариултууд хэрэглэгчдийг хангадаг болон чатбот хичнээн удаа хариулт өгөхийг зөвшөөрөх ёстой талаархи жишээнүүд орсон. Хүний харилцан үйлчлэлийн хүрэлцэх цэгүүдийг ойлгосны дараа, баг эдгээр ойлголтуудыг хэрэглэгчдэд илүү сайн туршлага хүргэх мэдээллийн урсгалын архитектурыг бүтээхэд ашигласан.
- Бот-хүн шилжилтийг удирдах:** Зарим тохиолдолд Jennie хангалттай хариулт өгч чадахгүй байж магадгүй. Ийм тохиолдолд, UX баг чатботыг хэрэглэгчийн хүсэлтийг болон чатны бүртгэлийг хэрэглэгчийн үйлчилгээний ажилтанд дамжуулахаас өмнө гурван удаа хариулах оролдлого хийхийг зөвшөөрсөн.

Ажилчид Jennie-тэй харилцах үед, MSD нь үндсэн хуудас дээр Jennie нь хиймэл оюун ухаан ашигласан бөгөөд цаг хугацааны явцад сайжрах бета хувилбар гэдгийг мэдээлэх болно (доорх зурагт үзүүлсэн) below).





## ДҮГНЭЛТ

3.25 Энэхүү тогтолцооны загварчлал нь бүрэн бус, төгсгөлгүй бөгөөд санал хүсэлт хүлээн авах нээлттэй баримт бичиг юм. AI технологиуд хөгжихийн хэрээр үүнтэй холбоотой ёс зүй болон удирдлагын асуудлууд ч мөн өөрчлөгдөнө. Хувийн мэдээллийг хамгаалах хорооны зорилго нь энэхүү загварчлалыг хүлээн авсан санал хүсэлтийн дагуу тодорхой хугацаанд шинэчилж, AI шийдэл ашиглаж буй байгууллагуудад үргэлжлүүлэн хэрэгцээтэй, ашигтай байлгах явдал юм.



*Хиймэл оюун ухааны хэрэглээний талаар зохистой харилцаа нь итгэлийг төрүүлж, байгууллага болон хувь хүмүүсийн хооронд нээлттэй харилцааг бий болгодог.*

## ХАВСРАЛТ А

### ОДООГИЙН ХИЙМЭЛ ОЮУН УХААНЫ ЁС ЗҮЙН ЗАРЧМУУДЫН ЭМХЭТГЭЛ

Энэхүү хавсралт нь янз бүрийн эх сурвалжаас шүүж авсан суурь AI-ийн ёс зүйн зарчмуудын эмхэтгэл юм.<sup>12</sup> Бүх сэдвийн дагуу нь тус загварчлалд оруулаагүй болно. Байгууллага эдгээр зарчмуудыг өөрсдийн компанийн зарчим, нөхцөл байдалдаа тохируулан шаардлагатай тохиолдолд тусгах талаар бодох нь зүйтэй юм.

1. **Хариуцлага:** Хиймэл оюун ухаанд оролцогчид өөрсдийн үүрэг, нөхцөл байдал, орчин үеийн дэвшилтэт технологид нийцүүлэн хиймэл оюун ухааны тогтолцооны хэвийн үйл ажиллагаа, хиймэл оюун ухааны ёс зүй, зарчмыг хүндэтгэн удирдах үүрэгтэй байх.
2. **Нарийвчлал:** Хүлээгдэж буй болон тохиолдох муу үр дагаврыг ойлгохын тулд алгоритм болон түүний өгөгдлийн эх сурвалжийн бүх алдаа, тодорхойгүй байдлын эх үүсвэрийг тодорхойлж, бүртгэж, тайлбарлаж, түүнийг бууруулах замаар мэдээлэх боломжтой.
3. **Шалгах чадвар:** Сонирхсон гуравдагч этгээдэд хяналт тавих, шалгах, шүүмжлэх боломжтой мэдээллийг нээн илрүүлэх замаар алгоритмын үйл ажиллагааг шалгах, ойлгох, хянах боломжийг олгодог.
4. **Тайлбарлах чадвар:** Автоматжуулсан болон алгоритмын шийдвэрүүд, тэдгээртэй холбоотой аливаа өгөгдлийг эцсийн хэрэглэгчид болон бусад сонирхогч талуудад техникийн бус үг хэллэгээр тайлбарлах боломжтой эсэхийг баталгаажуулах.
5. **Шударга байдал:**
  - a. Алгоритмын шийдвэр нь хүн амын өөр өөр нөхцөл байдалд ялгаварлан гадуурхах, шударга бус нөлөөлөл үүсгэхгүй байхыг баталгаажуулах.
  - b. Шийдвэр гаргах тогтолцоог хэрэгжүүлэхдээ санамсаргүй ялгаварлан гадуурхахаас сэргийлэх хяналт, бүртгэлийн механизмыг боловсруулах
  - c. Систем, программ, аппликейшн, алгоритмыг боловсруулахдаа олон төрлийн хүмүүсийн дуу хоолой, хөгжлийн зөвлөгөө авах.

<sup>12</sup> These include the Institute of Electrical and Electronics Engineers (“**IEEE**”) Standards Association’s *Ethically Aligned Design* (<https://standards.ieee.org/industry-connections/ec/ead-v1.html>), Software and Information Industry Association’s *Ethical Principles for Artificial Intelligence and Data Analytics* (<https://www.siiia.net/Portals/0/pdf/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990>) and Fairness, Accountability and Transparency in Machine Learning’s *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms* (<http://www.fatml.org/resources/principles-for-accountable-algorithms>). There is also the European Commission’s *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Building Trust in Human-Centric Artificial Intelligence* ([https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=58496](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58496)), and the OECD’s Recommendation of the Council on Artificial Intelligence (<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>). They also include principles raised through consultation feedback from the industry.

**6. Хүн төвтэй байдал болон Эрүүл мэндийн сайн сайхан байдал:**

- a. Мэдээллийн дадал зуршлуудын үр өгөөжийг шударга байдлаар хуваарилахыг зорьж, эмзэг бүлгийг хэт их хохироох мэдээллийн дадал зуршлуудаас зайлсхийх.
  - b. Мэдээлэл болон дэвшилтэт загварчлалын аргуудыг ашиглан хамгийн их үр өгөөжийг бий болгохыг зорих.
  - c. Хүний хөгжлийн, нэр төрийн болон бие даасан байдлын хувьд хувь нэмэр оруулах сайн сайхан ёс суртахуунуудыг дэмждэг мэдээллийн дадал зуршлуудыг хэрэгжүүлэх.
  - d. Мэдээллийн дадал зуршлуудад өртсөн хүмүүс эсвэл олон нийтийн сайн бодож тунгаасан шийдвэрүүдэд ач холбогдол өгч, тэдгээр хүмүүс эсвэл олон нийтийн үнэт зүйлс, ёс зүйн зарчимтай нийцүүлэх.
  - e. Шийдвэр гаргахдаа хувь хүнд урьдчилан таамаглахуйц хор хохирол учруулахгүй байхаар, эсвэл тийм хор хохирлыг хамгийн бага хэмжээнд байлгахуур байх.
  - f. Хэрэглэгчдэд ашиглагдаж буй мэдээллийг, түүнийг ашиглаж буй орчныг хянах, мөн тухайн ашиглалт болон орчныг өөрчлөх боломжийг олгох.
  - g. Хэрэглэгчийн нийт сайн сайхан байдлыг хиймэл оюун системийн үйл ажиллагааны үндсэн зорилго болгох.
7. **Хүний эрхийн зохицуулалт:** Технологийн загварчлал, хөгжүүлэлт, хэрэгжүүлэлт нь олон улсад хүлээн зөвшөөрөгдсөн хүний эрхийг зөрчихгүй байхыг баталгаажуулах.
8. **Хүртээмжтэй байдал:** Хиймэл оюун ухаан хүн бүрт хүртээмжтэй байхыг баталгаажуулах.
9. **Дэвшилтэт байдал:** Тухайн төсөлд оролцоогүй байснаас илүү үр өгөөжтэй үр дүнг бий болгодог хэрэгжүүлэлтүүдийг дэмжих

**10. Хариуцлага, хяналт, ил тод байдал:**

- a. Зохион бүтээгчид болон үйл ажиллагаа эрхлэгчид өөрсдийн систем, програм болон алгоритмуудад хариуцлага, хяналттай байж, эдгээр систем, програм болон алгоритмуудыг ил тод, шударга байдлаар ажиллуулахыг баталгаажуулснаар итгэлийг бий болгох.
- b. To make available externally visible and impartial avenues of redress for adverse individual or societal effects of an algorithmic decision system, and to designate a role to a person or office who is responsible for the timely remedy of such issues.
- c. Хиймэл оюун ухааны технологи хэрхэн, хэзээ хэрэглэгдэж байгааг хэрэглэгчид эсвэл оролцогч талууд шалгах боломжийг олгохын тулд урсгалын дараах арга хэмжээ, үйл явцыг нэвтрүүлэх.
- d. Загварчлалын процесс болон шийдвэр гаргалтын талаарх нарийвчилсан тэмдэглэл хөтлөх.

**11. Бат бөх байдал болон аюулгүй байдал:** Хиймэл оюун системүүд аюулгүй, хамгаалагдсан байж, сургагдсан өгөгдөлдөө халдлага болон зөрчилд өртөхөөргүй байх ёстой.

**12. Тогтвортой байдал:** Ирээдүйн үйл хөдлөлийг үр дүнтэй таамаглаж, тодорхой хугацаанд ашигтай мэдээлэл гарган авах боломжтой хэрэгжүүлэлтүүдийг дэмжинэ.

## ХАВСРАЛТ В

### АЛГОРИТМЫН АУДИТ

1. Алгоритмын аудит нь загваруудад орсон алгоритмуудын бодит үйл ажиллагааг илрүүлэх шаардлагатай үед хийгддэг. Энэ нь зохион байгуулалтын хариуцлагатай зохицуулагчаас (шүүхийн шинжилгээний хүрээнд) ирсэн хүсэлтийн дагуу эсвэл харилцагчийн байгууллагаа зохицуулагчын хүсэлтэд хариу өгөхөд туслах зорилгоор хиймэл оюуны технологийн нийлүүлэгчийн зүгээс хийгдэх ёстой. Алгоритмын аудит хийх нь техникийн мэдлэг шаарддаг бөгөөд гадаадын мэргэжилтнүүдийг татан оролцуулахыг шаарддаг байж болно. Аудитын тайлан ихэнх хүмүүс болон байгууллагуудын ойлгоход хүндрэлтэй байж болно. Алгоритмын аудит хийхэд шаардагдах зардал болон цаг хугацааг аудитын тайлангаас олж авах хүлээгдэж буй ашигтай харьцуулан үнэлэх ёстой. Эцэст нь, алгоритмын аудит нь шалгалтын хувьд тодорхой ашиг тустай байх нь тодорхой үед л хэрэгжүүлэх ёстой.
2. Алгоритмын аудитыг авч үзэхэд дараах хүчин зүйлс хамааралтай байж болно:
  - a. **Алгоритмын аудит хийх зорилго.** Моделийн хүрээ нь тайлбарлагдахуйц хиймэл оюун ухааны нэг хэсэг болох AI загварууд хэрхэн ажилладаг тухай мэдээллийг өгөхийг дэмждэг. Алгоритмын аудит хийхээс өмнө хувь хүмүүс, бусад байгууллагууд эсвэл бизнесүүд болон зохицуулагчид аль хэдийн олгосон мэдээлэл хангалттай, найдвартай эсэхийг авч үзэх нь зүйтэй.
  - b. **Аудитын үр дүнгийн зорилтот үзэгчид.** Энэ нь зорилтот үзэгчдийн өгөгдөл, алгоритм болон/эсвэл загваруудыг үр дүнтэй ойлгохын тулд шаардлагатай мэдлэг чадварыг хэлнэ. Өөр өөр үзэгчдэд шаардлагатай мэдээлэл харилцан адилгүй байдаг. Хэрэв үзэгчид хувь хүмүүсээс бүрдэж байвал, шийдвэр гаргах процесс болон хувь хүмүүсийн өгөгдлийг хэрхэн ашиглаж байгааг тайлбарлах нь тайлбарлагдахуйц хиймэл оюун ухааны зорилгыг илүү үр дүнтэй хэрэгжүүлэх болно. Харин үзэгчид зохицуулагчдаас бүрдэж байвал, өгөгдлийн хариуцлага болон алгоритмуудын ажиллагаатай холбоотой мэдээллийг эхлээд шалгах ёстой. Алгоритмын аудит нь хиймэл оюун ухааны загвар хэрхэн ажиллаж байгааг тодорхойлж чадах бөгөөд хэрэв түүний үйл ажиллагааны талаархи мэдээллийн үнэн зөв эсвэл бүрэн байдлыг эргэлзэх шалтгаан байвал үүнийг батлах боломжтой.
  - c. **Ерөнхий өгөгдлийн хариуцлага.** Байгууллагууд ерөнхий өгөгдлийн хариуцлагыг хэрхэн хангаж байгаа талаар мэдээлэл өгөх боломжтой. Үүнд Моделийн Хүрээний "Загвар Хөгжүүлэхэд зориулсан Өгөгдөл" хэсэгт дурдсан бүх сайн өгөгдлийн дадлууд багтана. Тухайлбал: Өгөгдлийн гарал үүслийн бүртгэлийг хөтлөх замаар өгөгдлийн удамшлыг хадгалах, Өгөгдлийн үнэн зөв байдлыг хангах, Өгөгдөлд агуулагдах туйлшралын хэмжээг багасгах, Өөр өөр зорилгоор өгөгдлийг хуваах, Өгөгдлийн үнэн зөвийг тодорхойлох, Өгөгдлийг тогтмол хянах, шинэчлэх.
  - d. **Арилжааны үнэ цэнэтэй мэдээлэл** Хиймэл оюуны загварууд дахь алгоритмууд зах зээлийн өрсөлдөх чадварт нөлөөлж чадах арилжааны үнэ цэнэтэй мэдээлэл байж болно. Жишээлбэл, алгоритм нь худалдааны нууц байж болох ба бизнесийн дүрмүүдийг агуулж байж болно. Хэрэв техникийн аудит хийхээр төлөвлөж байгаа бол үүнтэй холбоотой эрсдэлийг бууруулах арга хэмжээг мөн авч үзэх хэрэгтэй.



## ТАЛАРХАЛ

Хувь Хүний Өгөгдөл Хамгаалах Зөвлөл (PDPC) нь Моделийн Хүрээний талаар үнэтэй санал өгсөн дараах хувь хүмүүс болон байгууллагуудад чин сэтгэлийн талархлаа илэрхийлж байна (Үсгийн дарааллаар):

A*STAR	IBM Asia Pacific
Accenture	LawTech.Asia
AIG Asia Pacific Insurance Pte. Ltd.	Mastercard Microsoft
Apple	Asia
Asia Cloud Computing Association	MSD International GmbH (Singapore branch)
AsiaDPO	Non-Profit Working Group on AI
BSA   The Software Alliance	OCBC Bank
Cambrian AI	PwC
CUJO.AI	pymetrics
Data Synergies DBS	Salesforce
Element AI	Singtel
Emerging Technologies Policy Forum	Standard Chartered Bank
Facebook	Suade Labs
Fountain Court Chambers	Symphony AyasdiAI
Google	Telenor Group Temasek
Grab	International Tookitaki
Great Eastern	UCARE.AI
GSK	Untangle AI



## #SGDIGITAL

Singapore Digital (SG:D) нь Сингапурын дижиталчлалын хүчин чармайлтыг илэрхийлж, дижитал хөтөлбөрүүд болон санаачлагуудыг нэг дүрслэлээр тодорхойлон, дотоодын болон олон улсад ижил хэлээр хүргэж байна.

SG: лого нь улаан өнгийн илэрхийлэх цэгээс үүссэн дугуй хэлбэртэй фонтуудаас бүрддэг. SG нь Сингапурын товчлол бол D нь манай дижитал эдийн засгийг илэрхийлдэг. Инээмсэглэсэн дүрс нь Сингапурчуудын дижитал эдийн засагт шилжиж буй өөдрөг үзлийг илтгэдэг. Бид дижитал эдийн засагт урагшлах тусам бүх зүйл хүн төвтэй байх бөгөөд энэрэнгүй байдал, баталгаа нь бидний бүх үйл ажиллагааны гол цөм нь байх болно.



BROUGHT TO YOU BY

[www.imda.gov.sg](http://www.imda.gov.sg)

[www.pdpc.gov.sg](http://www.pdpc.gov.sg)

Copyright 2020 – Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC)

This publication is intended to foster responsible development and adoption of Artificial Intelligence. The contents herein are not intended to be an authoritative statement of the law or a substitute for legal or other professional advice. The PDPC and its members, officers and employees shall not be responsible for any inaccuracy, error or omission in this publication