

## **Алгоритмд суурилсан шийдвэрүүд: Ил тод, нээлттэй байдал ба байнгын хяналт**

*Катарина А. Цвайг*

- ▶ Алгоритм гэдэг нь математик хэлбэрээр тодорхойлогдсон асуудлыг шийдэх зарчмыг хэлнэ. Алгоритм нь алгоритмд суурилсан шийдвэр гаргах системийн (ADM систем) үндсэн нэг хэсэг бөгөөд үүнийг жишээлэхэд даатгалын салбарт эрсдлийг тодорхойлоход болон бусад салбарт ижил төстэй асуудлыг шийдвэрлэхэд ашигладаг.
- ▶ Эрсдлийн үнэлгээ хийхэд шийдвэр гаргах зарчим тодорхой бус байгаа бол алгоритмд суурилсан „бие даан суралцах чадвартай” байхаар программчлагдсан ADM системийг ашигладаг. Тус систем нь өнгөрсөн үед тохиолдсон үйл явдлууд, өгөгдлөөс суралцах, бие даан шийдвэр гаргах чадвартай гэсэн үг юм.
- ▶ Харин тухайн бие даан суралцах чадвартай системээс гаргасан аливаа нэг шийдвэр, үр дүн хэрхэн ямар дарааллаар гарч ирснийг тодорхойлох боломжгүй байдаг. Тийм ч учраас бие даан суралцах чадвартай ADM системийг дараах тохиолдолд онцгой хяналтад байлгах шаардлагатай. Үүнд: ард иргэдтэй холбоотой шийдвэр гаргахад ашиглагдаж буй, алдаатай шийдвэр гарсан тохиолдолд хувь хүн, нийгэмд асар их хохирол учруулж болзошгүй бол.
- ▶ Ингэхдээ зөвхөн тухайн алгоритмыг тусд нь авч үзэн хянах нь хангалттай бус. Үүнээс илүүтэйгээр тухайн ADM системийг (алгоритм болон өгөгдлүүд) бүхэлд нь, тус системийг ашиглаж буй нийгэм, мэдээлэл зүйн цогц системийг нийтэд нь хянах шаардлагатай.

## Агуулга

1. Алгоритмд суурилсан шийдвэр гаргах системүүд .....	2
2. Ил тод байдал, хяналт шалгалтын талаар өрнөж буй нийгмийн хэлэлцүүлэгт алгоритмд суурилсан шийдвэр гаргах системүүдийн аль хэлбэрийг нь илүү анхаарч үзэх шаардлагатай вэ? .....	6
3. Алдаа гарах эх сурвалж, алдаатай шийдвэр .....	7
4. Ямар тохиолдолд ADM системийг байнгын хяналтад байлгах шаардлагатай вэ? .....	9
5. Ил тод байлгах үүрэг хариуцлага .....	10
6. ADM системийн байнгын хяналт .....	11
7. Бие даан суралцах функц бүхий ADM системүүдийн эрсдлийн ангилалууд .....	15
8. Улс төрийн түвшинд авч болох арга хэмжээний сонголтууд .....	17
9. Ном зүй .....	19
Редакцийн зөвлөл .....	21

Дижиталчлал болон хиймэл оюун ухааны талаар олон нийтийн дунд өрнөж буй мэтгэлцээнд алгоритм гэсэн энэ ухагдахуун элбэг тохиолддог. Алгоритмыг төвөгтэй асуудлыг шийдэх ид шидийн шийдэл мэтээр хүмүүст ойлгуулах нэг хэсэг байна. Жишээлэхэд байгууллагын ажилтнуудын ажлын бүтээмжид үнэлгээ өгөх программ боловсруулж, зах зээлд нэвтрүүлсэн компани өөрийн бүтээгдэхүүнийг дараах байдлаар сурталчилж байх жишээтэй. Үүнд: „Эцсийн бүлэгт чанар сайтай мэдээллийн санг бий болгосноор аливаа зүйлийг урьдчилан таамаглах боломж хязгааргүй мэт болно” гэжээ. Нөгөө талаас алгоритм маш муу зүйл гэх ойлголт хүмүүст бий болсны улмаас алгоритмыг хянах байгууллагууд бий болгох, алгоритм үйлдлийг тайлбарлахыг шаардсан шаардлагууд олноор тавигдах болсон (Майэр-Шөйнбэргэр & Цүкиер, 2013 он). Алгоритмд хяналт тавихын тулд техникийн хяналтын алба үүсгэн байгуулахыг шаардаж байгаа хүмүүс, алгоритмыг ил тод болгохыг алгоритмын кодыг олон нийтэд нээлттэй болгох төдийхнөөр төсөөлөн ойлгож буй хүмүүс алгоритм гэсэн энэ ухаагдахуунуудыг хольж солин, олон янзын утгаар хэрэглэж байгаа нь харагдаж байна. Харин энэхүү асуудлыг нарийвчлан авч үзвэл алгоритм нь „алгоритмд суурилсан шийдвэр гаргах систем”-ийн нэгээхэн хэсэг нь юм. Алгоритмыг ганцаарчилан авч үзэж, хяналт тавих нь өрөөсгөл болохыг орчин үеийн судалгааны ажлууд харуулж байгаа. Үүнээс илүүтэйгээр алгоритмыг нийгмийн холбогдолтой процессуудад хэрхэн ашиглаж байгаад тогтмол хяналт тавих шаардлагатай. Яагаад ийм дүгнэлт гаргаж байна вэ гэдгийг ойлгомжтой болгохын тулд „алгоритмд суурилсан шийдвэр гаргах систем” болон „алгоритм” гэсэн энэ хоёр ухагдахууныг ялгах салгаж ойлгох нь чухал.

### 1. Алгоритмд суурилсан шийдвэр гаргах системүүд

Алгоритмд суурилсан шийдвэр гаргах системүүд (algorithmic decision making system буюу ADM систем гэж нэрлэдэг) нь шийдвэр гаргах явц, дарааллыг тодорхойлсон горимыг агуулсан байдаг. Энд жолоочийн тухай өгөгдөл мэдээлэлд тулгуурлан тээврийн хэрэгслийн даатгалын ангиллыг тогтооход хэрэглэгддэг энгийн системийг жишээ болгон дурдаж болох юм. Энэхүү системийн шийдвэр гаргах горим нь тодорхой, хүмүүст ойлгомжтой,

жолоочийн нас, урьд нь аваар осол гаргаж байсан түүх дээр ерөнхийдөө суурилдаг. Одоо үед нийгэмд өрнөж буй хэлэлцүүлэгт шийдвэр гаргах горимоо тухайн алгоритмын систем нь бие даан тодорхойлдог тогтолцоонууд нь маргааны үндсэн сэдэв болж байна.

### „Алгоритмус“ гэж үгийн гарал үүсэл

„Алгоритмус“ гэж энэ үг нь 9-р зуунд математик хэрэглээний талаар сурах бичиг бичсэн Ал-Кваризми хэмээх араб эрдэмтний нэрнээс гаралтай. Энэ үгийг логоритмус, ритмус гэх зэрэг үгстэй андуурах тохиолдол байдаг ч эдгээр үгс гарал үүслийн хувьд „алгоритмус“ гэж үгтэй ямар нэг ойролцоо утга, уялдаа холбоо байхгүй.

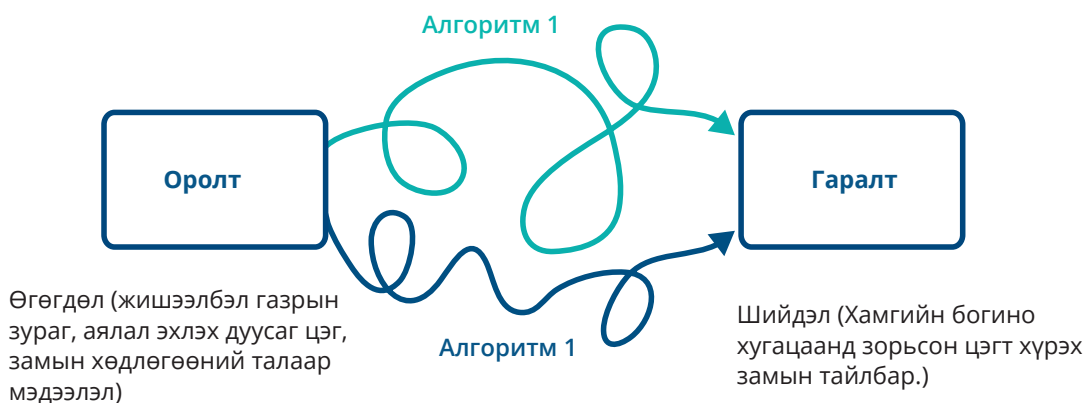
### Алгоритм гэж юу вэ?

Математик хэллэгт оруулж болох аливаа асуудалд алгоритмаар дамжуулан шийдэл олдог. Алгоритмыг төрөл бүрийн нөхцөл байдалд хэрэглэж болох бөгөөд тухайн нөхцөл байдалд тохирсон шийдлийг олох шаардлагатай болдог. Математик даалгавар нь ямар мэдээлэл байна вэ гэдгийг нэн түрүүн тодорхойлон (өгөгдөл), даалгаврын хариу ямар онцлог шинж чанартай байх шаардлагатайг тухайн өгөгдөл мэдээлэлд үндэслэн тодорхойлдог. Жишээлэхэд газрын зураг дээрх мэдээлэл, аялалын эхлэх, дуусах цэг, (хэрэв боломжтой бол) замын хөдөлгөний талаарх мэдээллийг ашиглан хамгийн богино замыг тооцох. Алгоритм нь ингэхдээ өгсөн мэдээлэлд (өгөгдлүүд) үндэслэн тодорхой нэг онцлог бүхий шийдэлд хэрхэн хүрэх гаргалгааг тодорхойлдог. Тодорхой нэг шийдэлд хүрэх алгоритм нь харин янз бүр байж болно (Зураг 1).

Математик  
даалгаврыг  
шийдэх  
гаргалгааны  
зарчим

### Математик даалгавар

Өгөгдөл болон шийдэл, тэдгээрийн хоорондын харьцааг тодорхойлдог.



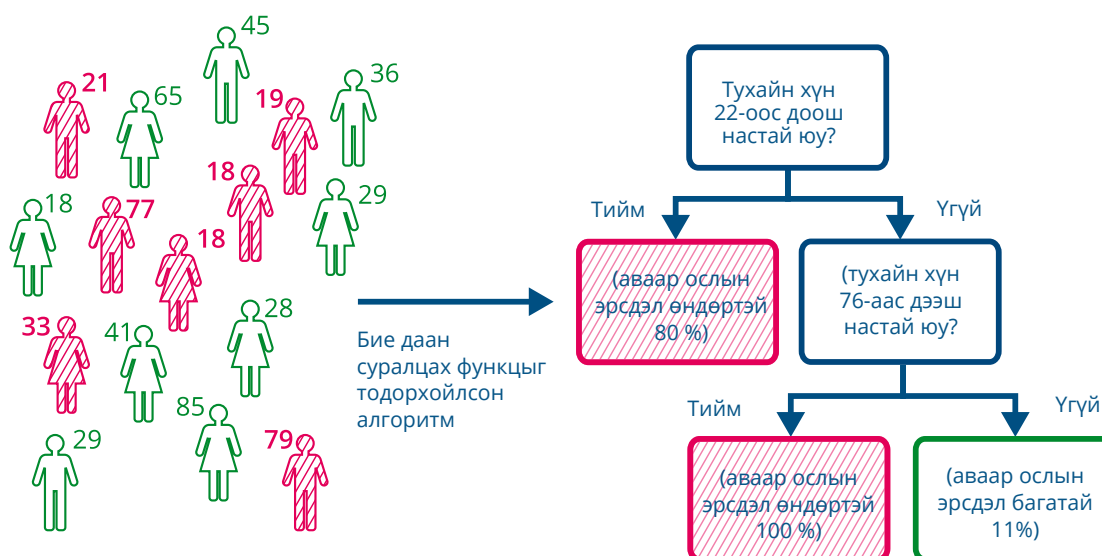
### Алгоритм нь

өгөгдлүүд дээр үндэслэн тодорхой нэг онцлог шинж чанар бүхий шийдэлд хэрхэн хүрч болох үйлдлүүдийн гаргалгаа буюу горим нь юм.

Зураг 1.

Математик асуудал, үүний шийдэл болох алгоритм гаргалгаа хоорондын харилцааг зургаар илэрхийлсэн байдал.

## Алгоритмд суурилсан шийдвэр гаргах системүүд



Зураг 2.

Алгоритмд суурилсан шийдвэр гаргах системүүдэд өгөгдөл өгч, тухайн өгөгдлийг ямар үйлдэлд суралцуулахаар зорьж байгааг тодорхойлон, ингэснээр тухайн систем нь шийдвэр гаргах горимыг өөрөө бие даан тодорхойлдог. Дээрх жишээн дээр жолоочийн нас, хүйс, сүүлийн 3 жил ямар нэг аваар осол гаргаж байсан талаарх (тийм бол улаан, үгүй бол ногоон) мэдээллийг өгсөн. Алгоритмын тусламжтайгаар жолооч нарыг нэг удаа осол гаргасан, эсвэл огт осол гаргаагүй гэсэн хоёр ангилалд хуваах горимыг тодорхойлсон. Энэхүү горимын тусламжтайгаар шинээр даатгалд хамрагдаж буй жолоочийн осол, аваар гаргах эрсдлийг тогтоож болно. Гэхдээ үр дүн нь өгөгдлөөс шууд хамааралтай гэдгийг энд анхаарах нь зүйтэй.

## Бие даан суралцах чадвар бүхий алгоритмуудын хэрэглээ

Өнөөгийн байдлаар нас, хүйс, санхүүгийн чадавх гэх мэт тодорхой өгөгдлүүд дээр үндэслэн ирээдүйд хэн зээл буцаан төлөх, хэн дахин гэмт хэрэг холбогдохыг найдвартай урьдчилан хэлэх систем бий болоогүй байна. Гэсэн хэдий ч хүн аливаа нэг зээлийг буцаан төлөх, эсвэл дахин хэрэг үйлдэхийг шийдэхдээ тодорхой зарчим баримталдаг нь тодорхой юм. Эсвэл хүмүүс тухайн нэг нөхцөл байдалд тодорхой нэг шийдвэр гаргах шалтгаан нь юу байна вэ гэдгийг тайлбарлах зүйлүүдийг тухайн хүний талаарх мэдээлэлд агуулагдаж байдаг гэж хэлж болох талтай. Бие даан суралцах чадвар бүхий алгоритмуудын тусламжтайгаар тухайн нэг хувь хүний мэдээлэл, тодорхой нэг нөхцөлд хийж буй үйлдлээс нь шийдвэр гаргаж буй зарчмыг нь тодорхойлох боломжтой юм (Зураг 2-ийг үзэх). Жишээлэхэд өөрөө бие даан суралцах чадвар бүхий алгоритмын системд жолоочийн нас, хүйс, тухайн хүн сүүлийн 3 жил аваар, осол гаргаж байсан эсэхийг оруулна. Үүний дараагаар алгоритм нь бие даан шийдвэр гаргах зарчмаа тодорхойлон шинээр бүртгүүлсэн жолоочийг ямар ч алдаагүйгээр „аваар осол гаргах өндөр эрсдэлтэй“, „аваар осол гаргах эрсдэл багатай“ гэсэн хоёр ангиллын аль нэгд нь хуваарилна. Тухайн алгоритм шийдвэр гаргах зарчмыг хэрхэн тодорхойлох нь „бие даан суралцах“ ямар арга барилыг хэрэглэснээс шалтгаална. Мөн шийдвэр гаргах

Мэдээлэл  
болон  
шийдвэр  
гаргах  
зарчмууд

зарчим нь харахад ойлгомжтой, тодорхой байх нь чухал. Харин алгоритмын өөрөө тодорхойлсон шийдвэр гаргах горим нь өгөгдөл мэдээллээс ихээхэн хамааралтайг энд тэмдэглэх нь зүйтэй.

## Яагаад „систем“ гэж нэрлэж байна вэ?

Алгоритмд суурилсан шийдвэр гаргах систем нь хоёр тус хэсэгт алгоритмыг агуулсан байдаг. Эхний алгоритм нь өгөгдлүүдийн тусламжтайгаар статистикийн моделийг тодорхойлно. Тус статистикийн модель нь хоёр дахь алгоритмын суурь болдог, ихэнх тохиолдолд маш хялбар маягаар тодорхойлогдсон байна. Энэхүү хоёр дахь алгоритм нь шинээр өгөгдөл өгөхөд хэрхэн шийдвэр гаргах вэ гэдэг горимыг тодорхойлдог. Зураг 2-т үндсэн шийдвэрийг тодорхойлсон алгоритм нь диаграмм хэлбэрээр зураглаж үзүүлсэн горимын дагуу тодорхойлогдсон. Гарах шийдвэр нь „аваар осол гаргах өндөр эрсдэлтэй“, „аваар осол гаргах эрсдэл багатай“ гэсэн энэ хоёр шийдвэрийн аль нэг байна. Алгоритмд суурилсан шийдвэр гаргах системээс гарах хариу нь өгөгдөл дата, түүнд дээр суурилан тооцоологдсон эхний алгоритм хоорондын харилцаагаар тодорхойлогдоно. Тэгэхээр өгөгдөл, тухайн системийн тооцоолсон нэгдүгээр алгоритм, тухайн систем шийдвэр гаргах үндсэн суурь болох хоёрдугаар алгоритм гэсэн энэ цогц системийг нийтэд нь авч үзэх шаардлагатай нь харагдаж байна.

Өгөгдөл мэдээллүүд, хоёр төрлийн хэрэглээний алгоритм

## Алгоритмын сонгодог хэрэглээ, орчин үеийн алгоритмд суурилсан шийдвэр гаргах системүүд

Шийдвэр гаргах зарчмыг (горимыг) өгөгдлүүд дээр үндэслэн тухайн систем өөрөө бие даан суралцан тодорхойлох арга барил нь алгоритмыг тодорхойлох сонгодог арга барилаас илүү үр дүнтэй, амжилттай байдаг. Компьютераар (алгоритм ашиглан) текст орчуулах үйл ажиллагааг жишээ болгон авч үзвэл үүнийг илүү ойлгомжтой тайлбарлах боломжтой. Энэ салбарт ямар нэг хэл дээрх текстийг өөр нэг хэл дээр орчуулахад зориулсан алгоритмын программчлал олон жилийн турш хийгдэж ирсэн. Энэ ажилд маш олон тооны чадварлаг хэл судлаачдыг татан оролцуулсан хэдий ч чанартай, сайн орчуулга хийх алгоритмыг бий болгож чадаагүй. Харин маш олон тооны текст, тэдгээрийн орчуулгыг дата өгөгдөл болгон оруулж, алгоритмын тусламжтайгаар шийдвэр гаргах зарчмыг бие даан тодорхойлох арга барилыг хэрэглэсэн нь амжилтын үндэс болсон байдаг.

„Бие даан суралцах чадвар“ бүхий алгоритмд суурилсан шийдвэр гаргах системүүд

Харин „бие даан суралцах чадвартай“ системийн шийдвэр гаргах горим хэрхэн тодорхойлогдож буй нь ил тод нээлттэй бус, түүнээс гадна өгөгдөл дата, алгоритм хоорондын болон тухайн системийг нийгэмд нэвтрүүлэхэд үүсэх харилцан уялдаа нь ойлгоход бэрх, төвөгтэй байдаг зэргээс шалтгаалан тухайн системээс гарсан шийдвэр нь хувь хүн, нийгэмд ихээхэн хэмжээний хохирол учруулж болзошгүй байдаг. Энэ тохиолдолд шийдвэр гаргах зарчим, үүнийг нийгмийн хэрэглээнд нэвтрүүлэхдээ чанарын хяналт тавих зайлшгүй шаардлагатай.

## 2. Ил тод байдал, хяналт шалгалтын талаар өрнөж буй нийгмийн хэлэлцүүлэгт алгоритмд суурилсан шийдвэр гаргах системүүдийн аль хэлбэрийг нь илүү анхаарч үзэх шаардлагатай вэ?

Алгоритмт суурилсан шийдвэр гаргах системүүд нь хүмүүсийн талаар, аливаа нэг хувь хүний эрх ашгийг хөндсөн эсвэл нийт нийгмийг бүхэлд хамарсан шийдвэр гаргаж буй бол тухайн системийг байнгын ажиглалт, онцгой хяналтад байлгах шаардлагатай. Аливаа нэг систем буруу шийдвэр гаргасан тохиолдолд тухайн системийг хэрэглээнд нэвтрүүлсэн этгээдээс гадна тухайн шийдвэр нөлөөлж буй хувь хүнд хор уршигтай. Байнгын ажиглалт, хяналтанд байх шаардлагатай алгоритмд суурилсан шийдвэр гаргах системүүдээс доор жишээ татъя. Үүнд:

Хяналт  
тавих өндөр  
шаардлагатай

1. Google-ийн хайлтын системийн алгоритмууд: Тус хайлтын систем нь вэб хуудас, хэрэглэгч, хэрэглэгчийн хандалт гэсэн мэдээллүүд дээр үндэслэн тухайн нэг илрэгц хайлтад ач холбогдолтой эсэхийг шийддэг.
2. Facebook newsfeed: Тус систем нь хэрэглэгч чухам ямар контентүүдийг хэрэглэж байна гэдгийг шинжлэн суралцаж, түүнд ямар мэдээлэл чухал байна гэдгийг шийдвэрлэн, дараа дараагийн утга агуулгыг шинээр харуулахдаа энэ суралцсан дээрээ үндэслэн горимоо шинээр тодорхойлдог.
3. Predictive policing (урьдчилан сэргийлэх стратеги гэсэн утгатай англи үг, орч.) гэдэг нь өнгөрсөн үеийн гэмт хэргүүдийн мэдээллүүдийг ашиглан гэмт хэрэг хэзээ, хаана гарч болзошгүйг тодорхойлох алгоритмууд дээр суурилсан байдаг. Германд иймэрхүү ADM системүүдийн программ хангамж хийгдэн, туршилтын шатанд явж байна.
4. Рецитив гэмт хэрэг гарч болохыг урьдчилан таамгалах алгоритмууд: АНУ-д гэмт хэрэгтнүүдийг дахин гэмт хэрэг үйлдэж болзошгүй байдлаар нь эрсдлийн бүлэгт хуваарилах ажлыг ADM системийн тусламжтайгаар гүйцэтгэдэг. Ийм төрлийн систем нь өнгөрсөн үеийн рецитив гэмт хэрэгтнүүд болон нийгэмшүүлэх хөтөлбөрт хамрагдсан этгээдүүдийн дата мэдээллээс суралцдаг, үүн дээрээ үндэслэн шийдвэр гаргах горимоо тодорхойлдог. Ийнхүү гарсан шийдвэрийг шүүх гэмт хэрэгтнийг дахин нийгэмшүүлэх хөтөлбөрт хамруулахыг шийдвэрлэхэд үндэслэл болгон хэрэглэдэг.
5. Террорист этгээдийг илрүүлэх үйл ажиллагаа: Snowden-ий олон нийтэд хүргэсэн бичиг баримтуудаас харахад террористүүдийн мэдээ хүргэгч байж болзошгүй этгээдийг илрүүлэхийн тулд ADM-тэй төстэй арга барилыг хэрэглэсэн байдаг. Афганистан, Пакистаны нийт 55 сая иргэний ухаалаг утасны мэдээлэл, харин террористүүдийн мэдээ хүргэгч цөөн тооны хүмүүсийг мэдээллийг үүнд ашигласан (Фойрч, 2015 он).

Энд нэмэж хэлэхэд нийгэмд буй нөөцийг хэрхэн ашиглах талаар буруу шийдвэр гаргаж буй, ингэснээрээ нийгэмд хохирол учруулж буй ADM систем байх магадлалтай. Энэ тохиолдолд ADM системийг нэвтрүүлж буй этгээдэд хохирол учруулахгүй байгаа хэдий ч нийгмийн хор уршиг өндөртэй байж болно. Жишээлэхэд тодорхой бараа бүтээгдэхүүн тээвэрлэхэд хамгийн хямд өртөгөөр тээвэрлэх замыг тодорхойлох чадвартай ADM систем байлаа гэхэд тус тээвэрлэлт нь харин байгаль орчинд хамгийн хор уршигтай байж болно. Энэ тохиолдолд ADM системүүдийг өөртөө хяналт тавьж чаддаг болгох шаардлагатай юу эсвэл урамшууллын механизмыг тэр чигээр нь өөрчлөх шаардлагатай юу гэдэг нь тодорхой бус байна. Мөн урамшууллын тогтолцоо нь угаасаа доголдолтой бол алгоритмын тусламжтайгаар нөөцийг маш ихээр үрэн таран хийж, улмаар нийгэмд асар их хэмжээний хохирол учруулах магадлал бий.

Алдаатай шийдвэр гарган, үүний улмаас хувь хүнд хохирол учруулж болох ADM системийг нэн түрүүн хяналтад авах шаардлагатай. Ийнхүү хяналт, шалгалт, ил тод нээлттэй байдлыг бий болгоход тавигдах шаардлагуудыг авч үзэхийн тулд эхлээд ямар тохиолдолд алдаатай шийдвэр гарч болохыг тайлбарлах нь зүйтэй.

Алдаатай шийдвэр – энэ нь хүмүүст нөлөөлж буй байдал

### 3. Алдаа гарах эх сурвалж, алдаатай шийдвэр

ADM систем алдаатай шийдвэр гаргахад хүрэх ерөнхийдөө 3 янзын механизм бий. Үүнд:

1. Санамсаргүй нөлөөлөх хүчин зүйлс: Аливаа нэг хүний ирээдүйн хөдөлмөрийн бүтээмж, суралцах чадамжийг урьдчилан таамаглах зорилго бүхий ADM системүүдэд („people analytics“ гэж нэрлэгддэг, Райндл, Крүйгл, 2017 он) санамсаргүй хүчин зүйлс, дижитал хэлбэрээр урьдаас тооцож болохооргүй зүйлсийн нөлөө их. Санамсаргүй хүчин зүйл гэдэг нь ажлын бүтээмж өндөртэй ажилтан өвчин тусах тохиолдол байж болно. Дижитал хэлбэрээр тооцож болшгүй зүйл гэдэгт ажлын байран дахь хүмүүс хоорондын харьцаа, багийн уур амьсгалыг нэрлэж болох юм. Эдгээр хүчин зүйлс нь ажлын бүтээмжид нөлөөлдөг. Энэ тохиолдолд шийдвэр алдаатай байх зайлшгүй нөхцөл бүрддэг.
2. Хэт бага дата өгөгдлүүд: Дата өгөгдлүүд өөр хоорондоо төстэй биш учраас байгаа өгөгдлүүдийг төстэй байдлаар нь хуваарилах боломжгүй. Жишээлэхэд нэг ижил шалтгааны улмаас нэг ижил шашны бүлэглэлд харъяалагдан, улмаар тэдэнтэй хамт террористууд үйл ажиллагаа явуулах хүний тоо маш цөөн учраас тэд нарыг автоматаар илрүүлэх боломжгүй. Энэ тохиолдолд мөн шийдвэр алдаагүй байх боломжгүй.
3. Боловсролуулалтын алдаа: ADM системийн буруу хэрэглээ. ADM системийн прогамм хангамжийг боловсруулахдаа алдаа гаргасан, эсвээс тухайн системээс гарсан үр дүнг буруу тайлбарласнаас

болж алдаатай шийдвэр гарч байгаа нь тогтоогдвол энэхүү алдааг боловсруулалтын явцад үнэлгээ явуулах тогтолцоог нэвтрүүлснээр багасгах боломжтой.

Дээр дурдсан алдаатай шийдвэр гарах нөхцлүүдийн эхний хоёр нь системээс үл шалтгаална. Тийм ч учраас дата өгөгдөл хэт бага байгаа бол ADM системийг ашиглахаас татгалзах шаардлагатай. Санамсаргүй хүчин зүйлийн хувьд тухайн нөхцөлд ADM системийг ашигласнаар гарч болох үр дүн, учирч болох хор уршиг нь юу байна гэдгийг харьцуулан авч үзсэний үндсэн дээр хэрэглэх эсэхийг шийдэж болох юм. ADM системийн алдаатай хэрэглээ, тухайн системийн боловсруулалтын явцад алдаа гарсан бол ил тод нээлттэй байдал, хяналт шалгалтын тусамжтайгаар энэхүү алдаанаас шалтгаалан гарч буй алдаатай шийдвэрийг багасгах боломжтой. Зарим нэг боломжуудыг доор жишээ зургаар үзүүлээ ( дэлгэрэнгүй тайлбарыг Цвэйг (2018 он) үз).

Алдааны эх сурвалжийг тогтоосноор яах вэ?

1. Мэдээллийн санд буй алдаа нь гарах шийдвэрт нөлөөлөх нь
  - а. Мэдээллийн санд мэдээлэл оруулахдаа зохистой мэдээлэл оруулах нь чухал. Жишээлэхэд COMPAS хэмээх рецитив гэмт хэрэг гарч болох эсэхийг урьдчилан таамгалах алгоритмын системд тухайн хүний үйлдсэн гэмт хэргийн талаархи мэдээлэллээс гадна тухайн хүний садан төрлийн холбоотой хүмүүс нь шоронд хоригдож байгаа эсэх мэдээллийг өгөгдлөөр оруулдаг. Германд алгоритмын ийм хэлбэрийг төсөөлөхийн аргагүй.
  - б. Өгөгдлүүд өөрөө алдаатай мэдээллийг агуулсан, эсвэл хэд хэдэн мэдээллийн санг хооронд нь холбоход тухайн нэг хүнд холбогдох мэдээллүүдийг найдвартайгаар тухайн хүний нэр дээр хуваарилах боломжгүй тохиолдлууд байж болно.
  - в. Өгөгдлүүд нь ямар нэг байдлаар ялгаварлан гадуурхсан шинжтэй байх нь бий. Жишээлэхэд ажилд орохоор материалаа өгсөн хүмүүсийн дунд дүрвэн ирсэн хүмүүс, эсвэл тийм хүмүүсийн ур удам, эмэгтэйчүүд материалаа өгсөн байлаа гэхэд эдгээр хүмүүсийн дундаас ярилцлаганд хамрагдсан хүмүүсийн тоо хангалттай хэмжээний бус байж болно. Энэ мэт ялгаварласан шинжтэй үйлдлийг алгоритм нэг ёсондоо өгөгдлүүдээс „өөрөө суралцсан“ байдаг. Харин ялгаварлан гадуурхалт гэж юу вэ, үүнийг хэрхэн тодорхойлох вэ гэдгийг тухайн нийгэм өөрсдөө шийдэх асуудал юм (Клайнбэрг, Муллайнатан & Рагаван, 2017 он; Цвайг & Краффт, 2018 он).
2. Алгоритм нь „өгөгдлүүдээс“ өөрөө „бие даан суралцаж, шийдвэр гаргаж буй энэ түвшинд алдаатай шийдвэр гаргах магадлалтай юм. Алгоритмыг бүхий л төрлийн асуудлыг шийдвэрлэхэд хэрэглэх нь тохиромжгүй.
3. Алгоритм нь өгөгдлүүдээс юуг суралцах шаардлагатай вэ гэдгийг нарийвчлан тодорхойлсон байх шаардлагатай. Жишээлэхэд социал медиагийн хүрээнд хэрэглэгдэж буй ADM системийн хувьд тухайн системийг боловсруулж буй дизайнерууд „relevance“ гэсэн энэ



ухагдахууныг тодорхойлсон программыг бий болгох шаардлагатай болдог. Энэ нь тодорхойлоход маш төвөгтэй нийгмийн дүр төрхийг агуулсан ухагдахууныг математик хэлбэрээр тодорхойлно гэсэн үг. Шийдэхэд төвөгтэй ч, шийдэж болох асуудал.

4. Ямар нэг компьютержсан үйл явцад хүн өөрөө дүн шинжилгээ хийх шаардлагатай гэж үзэж байгаа бол тухайн шийдвэр гаргах явцын дарааллууд нь тодорхой, гарч болох алдаануудыг тодорхойлсон байх нь чухал. Ингэснээр тухайн системд дүн шинжилгээ хийж буй хүн үр дүнг үнэлэх боломжтой болно.
5. Нийгмийн аливаа нэг тодорхой нөхцөл байдалд зориулан боловсруулсан ADM системийг, сайтар шалгалгүйгээр өөр төстэй нөхцөл байдалд зориулан хэрэглэх боломжгүй гэдэг нь өнөө үед тодохой болсон (Цвайг, Краффт 2018 он). Жишээлэхэд гэмт хэрэгтнүүдийг гэмт хэрэг давтан үйлдэх эрсдлийг урьчилан тооцох алгоритмын систем болох COMPAS нь гэмт хэрэгтнийг нийгэмшүүлэх хөтөлбөрийн хүрээнд арга хэмжээг хэрхэн хамгийн үр өгөөжтэй явуулах зорилгоор боловсруулагдсан. Одоогийн байдлаар тус системийг өөрчлөлгүйгээр хуулийн шийдвэр гаргахад ашиглаж байна. Эдгээр нь хоорондоо ялгаатай нийгмийн процессууд бөгөөд дээрх системийг шүүхийн шийдвэр гаргахад чанарын хувьд хангалттай бус гэдгийг судалгаагаар нотолсон. Шүүхийн шийдвэрт ашиглахын тулд чанарын өндөр стандартад тулгуурласан, мөн тухайн системийн шийдвэр гаргаж буй стандартуудыг өөр өнцгөөс авч үзсэн байх шаардлагатай юм (Краффт 2017 он).

Ингээд дүгнэж хэлэхэд ADM системүүдэд алдаа гарах магадлал өндөр, алдаа гарсан тохиолдолд хариуцах эзнийг тодорхойлоход хэцүү. Тийм учраас системийн чанарын стандартад өндөр хяналт тавьж, ил тод байлгахыг шаардаж байгаа нь үндэслэлтэй болох нь харагдаж байна.

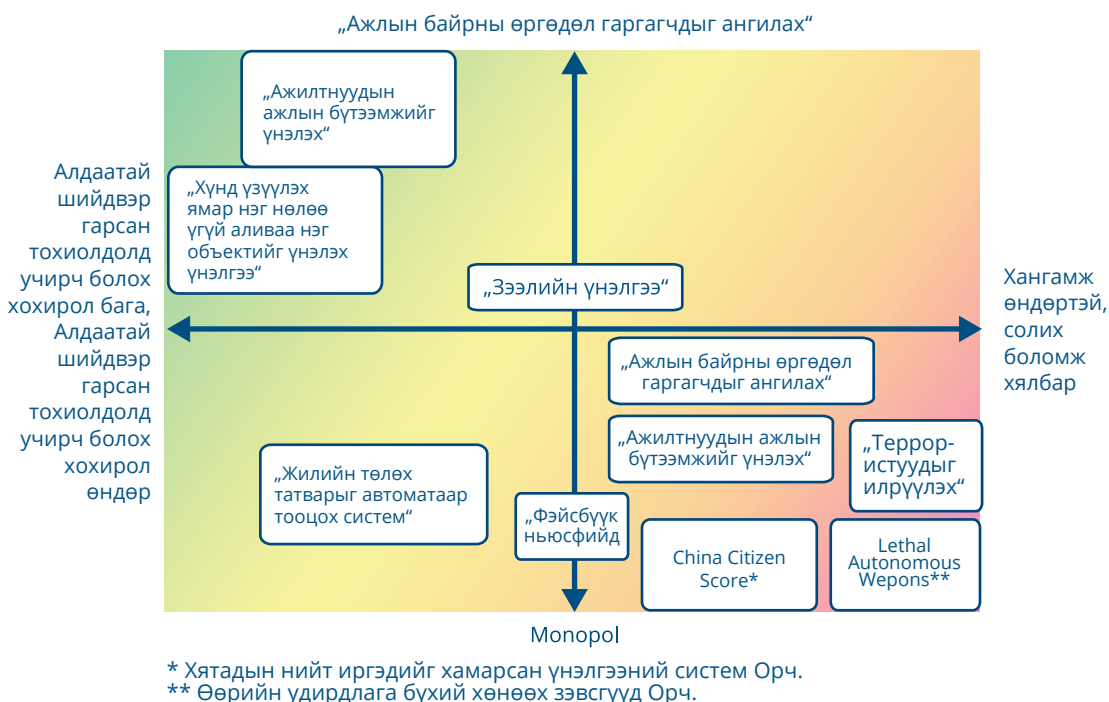
#### **4. Ямар тохиолдолд ADM системийг байнгын хяналтад байлгах шаардлагатай вэ?**

Хүмүүст үнэлгээ өгөх функц бүхий ADM систем алдаатай шийдвэр гаргаваас үр дагавар нь янз бүр. Зарим тохиолдолд тухайн хувь хүнд их хэмжээний хохирол учруулсан боловч хамрагдсан иргэдийн тоо цөөн байхад зарим системүүд хувь хүний түвшинд хор уршиг багатай ч нийгмийн хэмжээнд ихээхэн хохирол учруулдаг.

Учирч болзошгүй хохирол үнэхээр биеллээ олох магадлал буюу хохирлын үндсэн эрсдлийн магадлал нь хоёр дахь түвшний өөр нэг хэмжүүрээс ихээхэн хамаарна. Энэ нь хөндлөнгийн шинжээчийн үзэл бодол эсвэл хориг тавих арга хэмжээг хэр зэрэг яаралтай, хялбар хэрэгжүүлж болох вэ гэдгээс хамаарах юм.

Учирч  
болзошгүй  
хохирлын  
потенциал

„Энэхүү бүтээгдэхүүнийг худалдаж авсан хэрэглэгч, мөн доорх бүтээгдэхүүнийг худалдан авч байсан.“



Зураг 4. Зохиолчийн өөрийнх нь олж илрүүлсэн ADM системүүдийг харуулсан эрсдлийн матриц

Энэхүү хоёр хэмжээсийг харьцуулан үзсний үндсэн дээр тухайн үед шаардлагатай ил тод нээлттэй байдал, хөндлөнгийн хяналт, шалгалтын арга хэмжээнүүдийг шат дараалан хэрэгжүүлэх шаардлагатай.

## 5. Ил тод байлгах үүрэг хариуцлага

Алгоритмд суурилсан шийдвэр гаргах системүүдийг „ил тод“ байлгах ёстой гэсэн шаардлагууд байнга тавигддаг. „Ил тод байдал“ гэсэн энэхүү ухагдахууныг харин маш олон янзаар тайлбарлах нь бий. Жишээлэхэд тухайн систем „бие даан“ суралцахад шаардлагатай өгөгдлийн талаар ерөнхий мэдээлэл, өгөгдсөн мэдээлэл шийдвэр гаргахад хэдий хэмжээгээр нөлөөлж байгааг тайлбарласан тайлбар, алгоритмын кодыг олон нийтэд нээлттэй болгох гэсэн шаардлагуудыг энд дурдаж болно. Энэ шаардлагуудын дундаас хамгийн сүүлд дурдагдсан кодыг олон нийтэнд нээлттэй болгохыг шаардаж буй энэ шаардлага нь ил тод байдлыг бий болгох гэсэн энэ зорилгод тэр бүр нийцдэггүй. Учир нь алгоритмын кодыг роман мэт чөлөөтэй унших боломжгүй, нөгөөтэйгүүр программ хангамжийн компаниуд нь байгууллагын нууцлалыг хадгалахын тулд кодлолыг албаар маш төвөгтэй болгох боломжтой байдагт оршино. Google-ийн хайлтын системийн „Page Ranks“ буюу вэб хуудсыг эрэмбэлэхдээ хэрэглэж буй алгоритмыг олон нийтэд хүргэсэн нь вэб хуудсыг ажлуулж байгаа байгууллагууд олж авсан мэдээллээ овжиноор ашиглахад хүргэснээс бус нийгмийн зүгээс тавьж буй хяналт сайжирахад тэр бүр нөлөөлөөгүй. Зарим тохиолдолд „ил тод бус байсан“-аар арга заль хэрэглэхээс сэргийлэх боломжтой байх талтай.

Ил тод байдлыг бий болгох зохистой хэлбэрүүд

Үүнээс гадна зөвхөн кодыг хараад тухайн алгоритмд суурилсан систем хэрхэн ажиллаж байгааг хангалттай дүгнэх боломжгүй байдаг. Алдаа хаана гарч байна вэ гэдгийг тодорхойлохын тулд ямар дата өгөгдлүүдийг ашиглаж байна гэдгийг анхаарч үзэх нь мөн л төдий хэмжээний ач холбогдолтой. Зарим тохиолдолд системээс гаргасан шийдвэрийг амьдралд хэрэгжүүлэх тодорхой шийдвэр болгох үйл явц хэрхэн явагдаж байна гэдгийг ил тод болгох шаардлагатай байдаг. Жишээлэхэд ADM системийг рецитив гэмт хэрэг үйлдэх магадлалыг урьдчилан тооцоход хэрэглэж буйг дурдаж болно.

ADM системээс гарч буй алдаатай шийдвэр хэр зэрэг хор уршигтай вэ гэдгээс шалтгаалан ил тод нээлттэй болгох дараах арга хэмжээнүүдийг шат дараалан хэрэгжүүлэхийг үүрэг болгох нь зүйтэй. Үүнд:

1. Өгөгдөл мэдээллүүд болон тухайн ADM систем ямар шалгуурт суурилсан „бие даан суралцах“ функцтэй гэдгийг ойлгомжтой тайлбарлах. Ялангуяа хүмүүсийг тусгай категориудад ангилан, тэдгээрийн ирээдүйд хийж болох үйлдлийг таамгалах функц бүхий алгоритмын системд үүнийг хэрэгжүүлэх нь чухал.
2. Алгоритмын систем хэрхэн ажилладаг ерөнхий горим болон „бие даан“ суралцах функцууд хэрхэн ажилладгийг ил тод, ойлгомжтой тайлбарлах. Гэхдээ энэхүү ил тод, нээлттэй байдлыг бий болгох арга хэмжээг авснаар программын „бие даан“ суралцах чадварт томоохон хязгаарлалтыг бий болгоно гэдгийг анхаарах хэрэгтэй. Учир нь компьютерийг „өөрөө суралцах чадвартай“ болгон программчилснаар хүний ой тоонд багтахгааргүй, ойлгох боломжгүй үйлдлүүдийг хийх чадварыг бий болгодог. Энэ нь компьютерийн хамгийн чухал чадвар, энэхүү чадварыг хязгаарлана гэсэн үг юм.
3. Тусгай шинжээчдийн бага томилон тэдэнд ADM системийн талаар ойлголт өгөх.

Тухайн нэг систем алдаатай шийдвэр гаргаснаар учирч болзошгүй хохирлын хэмжээнээс шалтгаалан хяналтын янз бүрийн түвшний зохистой арга хэмжээг авч хэрэгжүүлэх шаардлагатай.

## 6. ADM системийн байнгын хяналт

Дээр дурдсан шалтгаануудын улмаас хяналт шалгалтын хүрээнд тухайн системийн „өөрөө бие даан“ суралцах алгоритмыг, эсвэл шийдвэр гаргаж буй үндсэн алгоритмыг мөн зөвхөн ADM системийг гэх мэтээр тусд нь ганцаарчилан авч үзэж үл болно. Хяналт, шалгалт нь ADM систем, тус системийг хэрэглэж буй болон тус системээс гарсан шийдвэр хүрэх хүн болгоныг хамарсан нийгмийн болон мэдээлэл зүйн нийт тогтолцоог бүхэлд нь авч үзсэн байх шаардлагатай (Зураг 4-ийг үз). Чанарын хяналтын стандарт арга хэмжээнүүд нь бие даасан системийн хэсэг тус бүрийг нэг бүрчлэн, мөн нийт тогтолцоог бүхэлд нь хамрах шаардлагатай.

Өгөгдөл дата-г  
ил тод, нээлттэй  
байлгах

Ил тод, нээлттэй  
байдлыг бий  
болгох алхамууд

Нийгэм,  
мэдээлэл зүйн  
цогц систем

Нийгэмд онц ач холбогдол бүхий ADM системийн программ хангамжийг боловсруулах үйл явцад хийх чанарын хяналтын стандарт нь ADM системийн найдвартай үйл ажиллагааг хангахад маш чухал үүрэгтэй. Чанарын хяналтын стандартууд нь жишээлбэл хэн, хаана, ямар шийдвэр гаргах, үүнийг хэрхэн баримтжуулах, үүнтэй уялдан чанарын хяналтын ямар арга хэмжээнүүдийг хэрхэн хэрэгжүүлэх зэргийг зохицуулсан байдаг. Үүнтэй адилаар нийгэмд чухал ач холбогдол бүхий ADM системийг нийгмийн аль нэг контекстэд нэвтрүүлэхдээ нийгэмд болж буй бусад процессуудтай хэрхэн уялдуулах тухай чанарын хяналтыг агуулсан стандартуудыг боловсруулсан байх шаардалгатай. Иймэрхүү стандартууд нь компьютерээр гарсан үр дүнг хэрхэн амьдралд шийдвэр болгон хэрэгжүүлэхэд зааварчилгаа өгөх, тухайн ADM систем нь шаардлагад нийцсэн эсэхийг шалгах зэрэг үйлдлүүд орно. Одоогоороо иймэрхүү тодорхой үе шаттай процесс байхгүй байсаар байна.

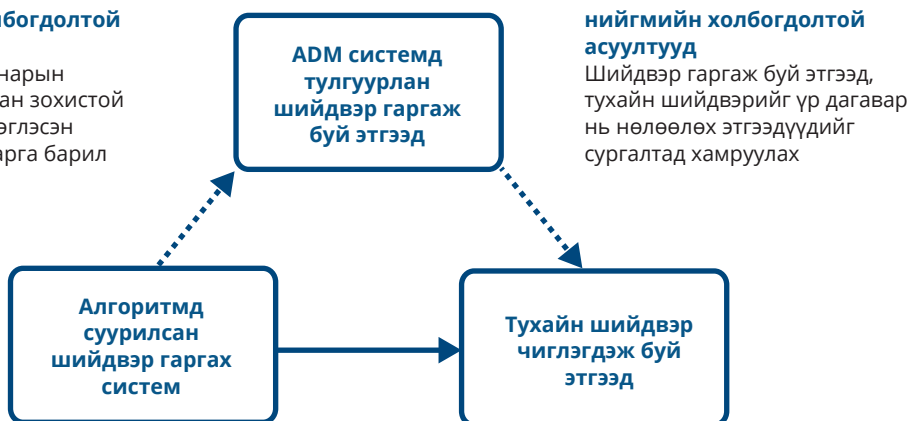
#### Техникийн холбогдолтой асуулууд

өгөгдөл дата чанарын стандарт хангасан зохистой алгоритмыг хэрэглэсэн дадлагуулах арга барил

ADM системд тулгуурлан шийдвэр гаргаж буй этгээд

#### нийгмийн холбогдолтой асуулууд

Шийдвэр гаргаж буй этгээд, тухайн шийдвэрийг үр дагавар нь нөлөөлөх этгээдүүдийг сургалтад хамруулах



#### Нийгэм, мэдээлэл зүйн цогц шинжилгээ

#### Нийгмийг хамарсан асуулууд

Жишээлбэл тухайн системийг гуравдагч этгээд өөрчилсөн хариуцлагыг хэн хүлээх вэ? ADM систем нь нийгэмд ямар өөрчлөлтийг бий болгож байна вэ?

Урамшууллын тогтолцоог өөрчлөх

#### Зураг 4.

ADM систем нь нийгэм дэх шийдвэр гаргах процесс хэв маягийг өөрчилдөг. Тухайн системийн функц хэрхэн ажилладагийг мөн хэр зэрэг нээлттэй болгосноос шалтгаалан ADM системийг тодорхой нэг зорилгоор ашиглаж болно. ADM системд хөндлөнгийн этгээд хязгаарлалт тавих боломжтой, жишээлбэл шүүх байгууллага. Шүүхийн энэ жишээг аваад үзвэл шүүхийн шийдвэр гарснаар, шүүхээс ял шийтгэл авсан хүн нь ямар нэг давж заалдах эрхтэй байх шаардалгатай. Тиймээс ADM системийн үйлдлийг хянан үнэлэх ажиллагаа нь тухайн систем хэрэглэгдэж буй нийгэм, мэдээлэл зүйн орчныг цогцоор авч үзсэн тохиолдолд л үр өгөөжтэй байна.

Ихэнх тохиолдолд ADM системийн үр дагаврыг тухайн систем хэрхэн ажиллаж буй функцийг нарийн мэдэхгүй байсан ч Black-Box-Analyse-ийн тусламжтайгаар хянах боломжтой байдаг (Крафт болон Цвайг, 2018 он). Аливаа ADM системээс учруулсан хохирол нь ихэнх тохиолдолд тухайн системээс гаргасан алдаатай шийдвэрийн хувь хэмжээнээс ихээхэн шалтгаалсан байдаг. Шийдвэр нөлөөлж буй этгээд ямар үйдлүүдийг хийж буйг ажиглах боломжтой бол тухайн системийн хэрхэн шийдвэр гаргасныг заавал мэдэх шаардлагагүйгээр системийн алдааг засах боломжтой. Жишээлэхэд сошиал медиа, компьютер хайлтууд дээр иймэрхүү ажиглалтуудыг явуулж болно. Эдгээр системүүдийн хайлтын үр дүнд харуулах мэдээ мэдээллийг автоматаар сонгодог, тухайн сонгосон мэдээлэл нь хүмүүсийн хайсан зүйлтэй нь таарч, хэрэглэгч сонгож байна уу гэдгийг ажиглах боломжтой. Ингэснээрээ алдаатай мэдээллүүдийн хувь хэмжээг хянах боломжтой.

black-box-анализ

---

Black-Box-Analyse-ийн тусламжтайгаар хянаж болох өөр нэг зүйл нь систем аливаа нэг ялгаварлан гадуурхсан шинжтэй шийдвэр гаргаж байгаа эсэх. (Сургалтын, мэргэжлийн) хүрээнд хүний бүтээмжийг үнэлэхэд алгоритмд суурилсан шийдвэр гаргах системийг ашиглаж буй бол юуны өмнө тухайн систем хүмүүсийг хүйс, гарал үүсэл, нас, бусад онцлогоор нь ялгаварласан шийдвэр гарч байгаа эсэхэд тогтмол хяналт тавьж байх шаардлагатай.

Ялгаварлан  
гадуурхалт үүсэх  
потенциал

---

Зарим тохиолдолд Black-Box-Analyse-ийн тусламжтайгаар ADM системийн онцлогоос шалтгаалсан асуудлуудыг шийдэх боломжтой. Алгоритмд суурилсан шийдвэр гаргах системүүд хувь хүний онцлогийг харуулсан өгөгдлүүдийг түлхүү авч үздэгээс filter bubble буюу тухайн хүний онцлогт зориулсан контентүүдийг хэт ихээр харуулах хандлага байдаг (Эли Парисэр 2011 он). Хэрэглэгчийн хайлтад filter bubble бий болсон байна уу гэдгийг арга барилын хувьд тодорхойлоход хэцүү ч хайлтын үр дүн тухайн хүний хувь хүний онцлогт хэт их тохирсон байна уу гэдгийг компьютерийн тусламжтайгаар харуулах боломжтой. Энэхүү чиглэлээр 2017 оны парламентын сонгуулийн үеэр Кайзэрслаутерн хотын Техникийн Их Сургуулийн Algorithm Accountability Lab-ээс төсөл хэрэгжүүлжээ. Харин энэ төслийн үр дүнгээр сайн дураар судалгаанд оролцсон иргэдийн интернэт хайлтын үр дүн нь тухайн хүний хувь онцлогт хэт чиглэгдсэн байдал байхгүй байна гэдгийг тогтоосон (Крафт болон бусад, 2017 он). filter bubble гэх мэт асуудал нь ихээхэн хэмжээний хохирол учруулж болзошгүй ч, „хэт хувь хүний онцлогт чиглэгдсэн” гэх мэт ганцхан онцлог чиглэлийг хянах шаардлагатай бол, энэхүү асуудалд хяналт тавих ажлыг томоохон хэмжээний хяналт, шалгалтыг бий болгохгүйгээр хямд төсөр өртөгтэйгөөр, байнга хэрэгжүүлээд байж болохыг энэхүү төслөөр харуулсан.

Хувь хүний  
онцлогт хэт  
тохирсон  
мэдээллүүд  
буюу filter  
bubble

---

Компьютер хүнтэй холбоотой шийдвэр гаргаж буй ихэнх тохиолдолд гарсан шийдвэрийн улмаас тухайн хүнд байж болох байсан боломжуудыг үгүй хийснийг, , нөгөө талаас тухайн хүн өөрт байх байсан боломжоо алдчихлаа гэдгийг нотлох боломж бас байдаггүй. Жишээлэхэд ажлын байрны танилцах уулзалтад уригдаагүй хүн өөрийгөө ажлыг сайн гүйцэтгэж чадах хүн гэдгийг

нотлох боломжгүй болно. Үүнтэй адилаар зээл олгохоос татгалзсан бол тухайн зээлийг хугацаанд нь төлж чадах байсан гэдгээ нотлох боломжгүй болох юм. Энэ тохиолдолд black-box арга барилыг ашиглах боломжгүй.

Аливаа нэг алгоритмд суурилсан шийдвэр гаргах системийг хэрэглэж буй хүн, түүнийг ажиллуулж буй хүмүүс өөрсдөө тухайн системээс гарсан шийдвэрүүдээс хамгийн сайн шийдвэр нь аль вэ гэдгийг тодорхойлох боломжгүй бол энэ нь асуудлыг үнэхээр төвөгтэй болгодог. Жишээлбэл сошиал медиагийн хүрээнд хэрэглэгдэж буй алгоритмд суурилсан шийдвэр гаргах системийн аль сонголт нь контентүүдийн хамгийн зөв сонголт болохыг тодорхойлоход хэцүү байдаг бол болзооны платформууд дээр хамгийн сайн зөв хосоо хэрхэн сонгох вэ гэдгийг тодорхойлоход мөн л төвөгтэй юм.

Дээр дурдсан ADM системийн алдаатай шийдвэрүүдийн улмаас гарч болох хор уршигийн улмаас ADM системүүдийг, нэн ялангуяа АНУ-ын тагнуулын байгууллагын ашигладаг террористуудад хэл хүргэгч нарыг олж илрүүлэхэд зориулсан ADM системүүдийг байнгын, эрчимжүүлсэн хяналтад байлгах шаардлагатай. Эдгээр систем нь ямар нэг сэжигтэн этгээд илрүүлэн, тухайн этгээд нь буруугүй тохиолдолд хувь хүнд учрах хор хохирлын хэмжээ асар их байх болно. Нөгөө талаас террористүүдийг олж илрүүлж чадахгүй бол нийгэмд учирч болзошгүй хохирол нь маш өндөр. Энэ системд үндэслэн шийтгэл оноосон хүмүүсийн тоо маш өндөр, үүнтэй зэрэгцэн байнга өндөр хяналттай байдаг нь зайлшгүй. Иймэрхүү ADM системүүдийг нээлттэй (жишээлэхэд магадлан итгэмжлэгдсэн байгуулагаас, эсвэл бие даасан, хараат бус шинжээчдийн багаар хянуулж байх) байлгаж, нийт нийгэмд үзүүлж болох нөлөөны цар хэмжээг тогтмол хянаж байх нь зүйтэй. Үүнтэй адилаар дайны талбарт андууран өөрийн зэвсэгт хүчнийг сөнөөхөөс сэргийлэх, татварын тайлангийн алдаа, зөрчлийг хянах, VISA картыг автоматаар хянах зэрэг үйл ажиллагаануудад ADM системүүдийг хэрэглэж буй бол, хяналтад онцгой анхаарах шаардлагатай.

Нийгмийн ач холбогдол бүхий ADM системүүдийн эрсдлийн аль ангилалд хамруулан, дээр танилцуулсан ил тод, нээлттэй байлгах үүрэг хариуцлага, техникийн хяналтын процессуудыг хэрэгжүүлэх шаардлагатай байна гэдгийг тодорхойлж болно.

## 7. Бие даан суралцах функц бүхий ADM системүүдийн эрсдлийн ангиллууд

Нийгмийн ач холбогдол бүхий ADM системүүдийг дараах таван ангилалд хувааж авч үзэхийг миний зүгээс санал болгож байна. Үүнд:

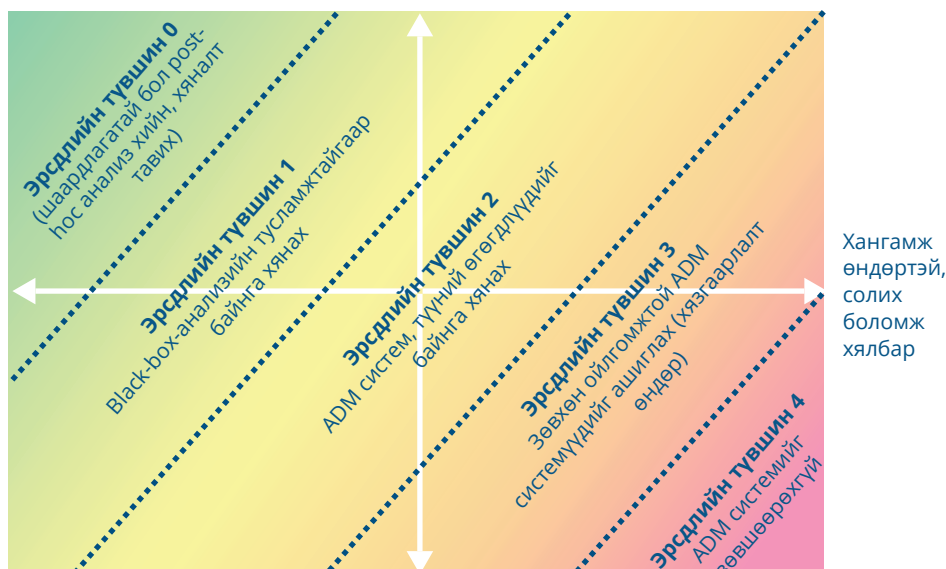
- ▶ Эрсдлийн 0 ангилалд хамрах ADM системүүдэд техникийн түвшинд зохицуулалт хийх ямарваа нэг шаардлага байхгүй. Тийм ч учраас ил тод, нээлттэй байлгах үүрэг хариуцлага хүлээлгэхгүй, байнгын хяналтад байлгах шаардлагагүй. Сэжигтэй тохиолдол илэрсэн бол post-hoc-анализ (аливаа нэг тохиолдол болж өнгөрснөөс хойш хийгдэж буй дүн шинжилгээ, орч.) хийж болох юм. Аливаа нэг сэжигтэй тохиолдол илэрснээр тухайн системийн нийт хохирлын потенциал өндөрсөж, улмаар тухайн системийг цаашид өндөр эрсдэлтэй дараагийн ангилалд авч үзэн хяналтыг чангатгах шалтгаан болно. Энэхүү ангилалд жишээлэхэд бэлэн бүтээгдэхүүнийг санал болгох системийг хамруулж болно.
- ▶ Эрсдлийн 1-р ангилалд хамрах ADM системүүдийн кодыг нээлттэй болголгүйгээр, black-box-анализийн тусламжтайгаар байнга хяналтад байлгахыг санал болгож байна. Миний бодлоор энэхүү ангилалд google-ийн автомат хайлт хамрагдана. 2017 онд явуулсан судалгаагаар google-ийн автомат хайлтын системд хэрэглэгдэж буй алгоритмын систем нь хувь хүний онцлогийг харуулсан өгөгдлүүдийн түвшин маш доогуур учраас filter bubble үүсэх магадлал багатай гэдгийг баталсан. Black-box-анализийг байнга явуулж байх шаардлагатай бөгөөд хувь хүний онцлогийг харуулсан өгөгдлийн түвшин өндөрсөн тохиолдолд учруулж болох хор хохиролд шинээр үнэлгээ хийх шаардлага гарч ирнэ.
- ▶ Эрсдлийн 2-р ангилалд хамрах ADM системүүд нь ноцтой тогтолцоонууд учир байнгын хяналтаас гадна, ил тод нээлттэй байлгах үүрэг хариуцлага хүлээлгэх шаардлагатай. Үүнд: Системийн өгөгдөл дата-нуудын талаарх нарийн мэдээлэл, өгөгдлүүдэд тавигдаж буй чанарын стандарт, түүний шалгуурууд, шударга байдлын зохистой шалгуурууд, мөн эцсийн шийдвэр гаргах нийгмийн процессийг тухайн системийн хувьд хэрхэн нийгмийн бусад процессуудтай уялдуулах зэрэг багтана. Энэ ангилалд жишээлбэл ажлын байрны зараар өргөдөл гаргасан хүмүүсийн өргөдлийг үнэлэх ADM-ийн автомат буюу туслах системүүд хамрагдана. Энэхүү системүүдийн хувьд өгөгдлийн аль аль мэдээллүүдийг ашиглаж болох, хүргүүлсэн материалуудыг чанарын ямар стандартыг ашиглан сонгох, тухайн ADM систем нь ямар шалгуурын дагуу бэлтгэгдсэн зэргийг нарийн тодорхойлох шаардлагатай. Амжилтгүй болсон ярилцлаганы тоо байж болох хамгийн доод түвшинд барих, эсвэл ярилцлага нь амжилттай болж, ажлын гэрээ байгуулсан хүмүүсийн тоог байж болох хамгийн дээд хэмжээнд хүргэхийн тулд „амжилттай” гэдэг энэ ухагдахууныг

„ажлын гэрээ байгуулснаас хойш байгууллагад хоёр жил тасралтгүй ажилласан гэж“ тодорхойлсон байж болно. Энэ тохиолдолд жирэмсний амралт авсан хүмүүст энэ шалгуур нь сөргөөр нөлөөлнө. Эцсийн эцэст шийдвэрийг бүрэн автоматжуулсан хэлбэрээр гаргасан уу, эсвэл зөвхөн шийдвэр гаргах бэлтгэл ажилд ашигласан уу гэдгийг, мөн шийдвэрийг эсэргүүцэн даваж заалдах ямар боломжууд байгааг ил тод байлгах нь нэн чухал юм.

- ▶ Эрсдлийн 3-р ангилалд хамрагдах ADM системүүдийн „бие даан суралцагч“ бүрэлдэхүүн хэсэг нь зөвхөн тайлбарлаж болох моделуудийг ашиглах шаардлагатай. Дээр харуулсан шийдвэр гаргах процессийг тайлбарласан диаграммыг тайлбарлах боломжтой систем гэж үзнэ. Учир нь энэхүү системээс гарч буй шийдвэр нь ойлгомжтой, юунаас хамааран тухайн шийдвэр гарч буй нь тодорхой. Өнөө үед deep learning-ийн талбар дахь (хиймэл) мэдрэлийн системийн сүлжээнүүдэд хэрэглэгдэж буй моделуудыг тайлбарлах боломжгүй моделиуд гэж болно. Энэ ангилалд аливаа нэг байгууллагын ажилтнуудыг ирээдүйд үзүүлэх амжилтаар эрэмбэлэн ангилах функц бүхий системууд хамрагдана.
- ▶ Эрсдлийн 4-р ангилалд миний бодлоор „бие даан суралцах чадвартай“ нэгж хэсгийг агуулсан алгоритмын тусламжтайгаар шийдвэр гаргахыг зөвшөөрч үл болох нөхцлүүдийг хамруулах шаардлагатай. Жишээлбэл эрэн сурвалжилж буй этгээдийг хөнөөх шийдвэрийг автоматаар гаргадаг зэвсгүүд (lethal autonomous weapons), иргэдийн үйл хөдлөлд үнэлгээ өгөх нийтийг хамарсан хяналтын систем (China Citizen Score), террористуудыг илрүүлэх системүүд (Skynet) гэх мэт.

Дээр дурдсан эрсдлийн 5 ангиллыг 5-р зурагт үзүүлэв.

Алдаатай шийдвэр гарсан тохиолдолд учирч болох хохирол бага, Алдаатай шийдвэр гарсан тохиолдолд учирч болох хохирол өндөр



Зураг 5.

Нийгмийн ач холбогдол бүхий ADM системүүдийг эрсдлийн матрицийн тусламжтайгаар эрсдлийн таван ангилалд хуваах нь



## 8. Улс төрийн түвшинд авч болох арга хэмжээний сонголтууд

Улс төрийн түвшинд хэрэгжүүлж болох олон тооны арга хэлбэрүүд байдаг. Эдгээрийг доор авч үзье.

### 8.1. Алгоритмд суурилсан шийдвэр гаргах тогтолцоонуудыг хянах, нээлттэй болгох тухай нийгмийг хамарсан хэлэлцүүлгүүд

Герман улс аж үйлдвэр маш өндөр хөгжсөн орон. Тийм учраас аливаа нэг зохицуулалтын арга хэмжээг үнэхээр шаардлагатай тохиолдолд л авч хэрэгжүүлдэг. Тиймээс алгоритмд суурилсан шийдвэр гаргах систем болгоныг техникийн хяналтад оруулах шаардлагагүй байдаг (алгоритм болгоныг хянах шаардлага бол бүр байхгүй гэсэн үг). Харин өнөө үед иргэдийн анхаарлыг маш их татаж буй сошиал медиад хэрэглэгддэг алгоритмд суурилсан системүүдэд л зөвхөн анхаарлаа хандуулах нь бас өрөөсгөл юм. Учруулж болох хор хохирлыг аваад үзвэл улсын байгууллагуудад хэрэглэгдэж буй ADM системүүдэд, хувийн байгууллагад ажлын байрны өргөдөл гаргасан хүмүүс, байгууллагын ажилтнуудыг үнэлэх үнэлгээнд хэрэглэгдэж буй ADM системүүдийн нэг адил хяналт шаардлагатай. Энд дурдахгүй байхын аргагүй нэг зүйл бол lethal autonomous weapons буюу автомат зэвсгийн системүүд. Тиймээс нийгмийн ямар нөхцөл байдалд хэрэглэгдэж буй ADM системийн аль хэлбэрт нь ямар хэлбэрийн хяналт, шалгалт хэрэгтэй талаар нийгмийг хамарсан, үр өгөөжтэй, конструктив хэлэлцүүлэг өрнүүлэх шаардлагатай байна. Энэ судалгааны хүрээнд ADM системийг эрсдлийн хувьд хэрхэн ангилах талаар санал дэвшүүлсэн байгаа.

### 8.2. Программ хангамжийн болон нийгэмд нэвтрүүлэх үйл ажиллагааг тодорхой бүтцэд оруулах нь

Нийгэмд болон мэдээлэл зүйн хувьд ач холбогдол бүхий ADM системийн программ хангамжийг боловсруулахад тогтсон стандартууд шаардлагатай байна. Үүний зэрэгцээ ADM системийг нийгэмд хэрэгжүүлэхэд тодорхой стандартууд мөрдөгдөх шаардлагатай. Үүнд хэрэглэгч нарыг сургалтанд хамруулах, гарч буй шийдвэрт чанарын хяналт тавих гэх мэт үйл ажиллагаанууд орно. Холбооны болон мужийн түвшинд дэх улсын байгууллагууд юуны өмнө ADM системийг агуулсан мэдээллийн тоног төхөөрөмж худалдан авах, тендер зарлахдаа тодорхой дэс дараалал бүхий үйл ажиллагааны стандартуудыг бий болгон, тухайн стандартыг бүтээгдэхүүн худалдаалж буй байгууллагаас дагаж мөрдөхийг шаардан, тухайн программыг өөрсдийн нийгмийн чанартай үйл ажиллагаанд хэрэглэх үйл явцыг тодорхой дэс дараалалтайгаар хэрэгжүүлэн, нийт нийгэмд үзүүлж буй үр нөлөөг нь үнэлэх үнэлгээ байнга хийж байвал энэхүү стандартуудыг ажил хэрэг болгоход томоохон хувь нэмэр болох юм.

### 8.3. Мэргэжлийн сургалт, мэргэжлийн дүр төрх, мэргэжлийн ёс зүй

Дээр дурдсан үйл явцад шаардлагатай асуулт тавьж чадах, үйл явцад чанарын түвшний үнэлгээ өгч чадах мэргэжлийн боловсон хүчин мөн шаардлагатай байна. Кайзэрслаутерн хотын Техникийн их сургуульд (одоогийн байдлаар Герман орон даяар зөвхөн энд) нийгмийн мэдээлэл зүй гэсэн чиглэлээр сурах боломжтой байна.

Үүнээс гадна Data scientist гэсэн мэргэжлийн хүмүүс байгаа бөгөөд эдгээр нь ADM системийг программ хангамжийг боловсруулахад оролцдог. Энэхүү мэргэжил нь шинэ учир суралцах хичээлийн нэгдмэл төлөвлөгөө, мэргэжлийн ёс зүй зэрэг нь тодорхойлогдоогүй байна. Data scientist чиглэлээр ажиллаж буй ихэнх хүмүүс нь физик, математик, мэдээлэл зүйн чиглэлээр сурсан, статистик, програмчлалын өргөн хэмжээний мэдлэгтэй хүмүүс байх боловч, ихэнх тохиолдолд нийгмийн ухааны чиглэлээр ямар нэг сургалтад хамрагдаагүй хүмүүс байдаг. Тиймээс энэ чиглэлээр тэнхим байгуулах ажлыг дэмжих, хичээлийн тогтсон төлөвлөгөө боловсруулах нь чухал байна.

Үүнээс гадна Германд мэдээлэл зүйн чиглэлээр суралцаж болох бүхий л их, дээд сургуулиуд дээр хуучин байсан хэдий ч одоо бараг байхгүй болчихсон байгаа „Мэдээлэл зүй ба нийгэм“ чиглэлээр тэнхимүүдийг дахин нээх шаардлагатай байна.

Эцэст нь хэлэхэд дээрх чиглэлээр мэргэжлийн ёс зүйг боловсруулах, нийгэм, мэдээлэл зүйн ач холбогдол бүхий ADM системийг хянах үүрэгтэй албан тушаалтнуудыг магадлан итгэмжлэлд хамруулах шаардлага байна. Ингэснээрээ энэ нь байгууллагын нууц алдагдах вий хэмээн санаа зовниж буй хувийн байгууллагуудын асуудлыг авч үзэх шийдэл болно.

## 9. Ном зүй

Михайл Фөйрч, NSA-ийн хяналт, Skynet гэж байдаг, үнэхээр аймшигтай, 2015 оны 5 сарын 11-ний өдөр онлайн нийтэлсэн, <https://www.wired.de/collection/tech/das-nsa-program-skynet-soll-terroristen-identifizieren>, (2018 оны 11 сарын 12-ны өдрийн хандалт).

Жон Клайнбэрг, Сэндил Муллайнатан & Маниш Рагаван, эрсдлийн үнэлгээг тэгш тодорхойлоход гарч буй зөрчлүүд, „Компьютерийн онолын шинжлэх ухаан дах инновацууд“ сэдэвт 8-р бага хурал (ITCS'17), 2017 он, 43:1 – 43:23.

Тобиас Д. Краффт, Катарина А. Цвайг; Нийгмийн зүгээс алгоритмд суурилсан шийдвэрүүдийг хэрхэн нарийвчлан шалгаж болох вэ?, „Тооцоолох боломжтой юу (бус уу)? Нийгэм, улс төрийн түвшин дэх автоматжуулалт, алогоритмийн хэрэглээ“ ном, Эрхлэн нийтлүүлсэн (Реза Мохаббат Кар, Басанта Тапа, Петер Паричек), „Нээлттэй мэдээлэл зүй“ Чадавхийг сайжруулах төв, 2018 он, 471-492.

Тобиас Д. Краффт, Михайл Гамер, Марсэл Лайссинг, Катарина А. Цвайг, Хөөс хагарсан уу? Google-ээр Бундэстагийн сонгуулийн тухай хайхад хувь хүний онцлог хэт тохирсон илрэл маш бага, „мэдээллийн сангийн хандив“ – төслийн дунд шатны тайлан, 2017 он, [https://algorithmwatch.org/wp-content/uploads/2017/09/1\\_Zwischenbericht\\_final.pdf](https://algorithmwatch.org/wp-content/uploads/2017/09/1_Zwischenbericht_final.pdf) (2018 оны 11 сарын 12-ны өдрийн хандалт).

Тобиас Д. Краффт, Гэмт хэргийг урьдчилан таамаглах хэрэгслүүдийг ашиглаж буй ажлын хүрээнд хоёртын системд суурилсан ангилалт хийх зарчмуудын чанарын түвшин, Кайзерслаутерн хотын Техникийн Их сургууль, мастерийн ажил, 2017 он, <https://arxiv.org/abs/1804.01557> (2018 оны 11 сарын 12-ны өдөр хамгийн сүүлд хандав).

Конрад Лишка, Анита Клингел: Кодпьютер хүний талаар шийдвэр гаргах тохиолдолд – алгоритмд суурилсан шийдвэр гаргах үйл явцыг олон улсын жишээн дээр авч үзсэн нь, Бертелсманн Сангаас явуулсан судалгаа, 2017 оны 5 сар, [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/ADM\\_Fallstudien.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/ADM_Fallstudien.pdf) (2018 оны 11 сарын 18-нд хамгийн сүүлд хандав).

Виктор Майер-Шөйнбергер, Кеннет Цукиер: Big Data: Бидний амьдралыг өөрчлөх хувьсгал, Редлайн хэвлэлийн газар, Мюнхен хот, 2013 он.

Эли Парисэр: Filter Bubble: бид интернэтийн сүлжээнд хэрхэн амаа үдүүлж байна вэ, Карл Ханзер хэвлэлийн газар, Мюнхен хот, 2011 он.

Корнелиа Райндл, Штефаний Крүйгл: Peopla Analytics-ийн практик хэрэглээ, Haufe-Lexware GmbH & Co. KG., Фрайберг хот, 2017 он.

Катарина А. Цвайг, Конрад Лишка, Др. Сара Фишер: Компьютер хаана алдаа гаргаж байна, Бертелсманн Сангаас явуулсан судалгаа, 2018 оны 2 сар, <https://www.bertelsmann-stiftung.de/fileadmin/>

[files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf](#) (2018 оны 11 сарын 12-ны өдөр хамгийн сүүлд хандав).

Катарина А. Цвайг, Тобиас Б. Краффт: Алгоритмд суурилан гарч буй шийдвэрийн чанар, шударга байдал, „Тооцоолох боломжтой юу (бус уу)? Нийгэм, улс төрийн түвшин дэх автоматжуулалт, алогоритмийн хэрэглээ“ ном, Эрхлэн нийтлүүлсэн (Реза Мохаббат Кар, Басанта Тапа, Петер Паричек), „Нээлттэй мэдээлэл зүй“ Чадавхийг сайжруулах төв, 2018 он, 204-227.

## Редакцийн зөвлөл

### Зохиолч

Проф. Др. Катарина А. Цвайг

Кайзерслаутерн хотын Техникийн их сургуулийн Мэдээлэл зүйн тэнхимийн харъяа Alogorithm Accountabilty Lab-ийн захирал. Тэрээр тус их сургуулийн, Герман орон даяар ганц байгаа „Нийгмийн мэдээлэл зүй“ тэнхимийн зохицуулагч. Энэхүү тэнхим нь комьютерийн программ, байгууллага, хувь хүн, нийгэм хоорондын харилцан уялдаа, нөлөөллийг загварчлах, түүнд дүн шинжилгээ хийх, боломжтой бол бүрэн хэмжээгээр загварчлах ажлыг гүйцэтгэдэг. Проф. Др. Катарина А. Цвайг нь олон тооны шагнал, урамшуулалд хамрагдаж байсан (2017 оны Мэдээлэл зүй, инженерийн шинжлэх ухааны салбарын тэнхимийн багш нарт олгодог ars legendi шагнал, Algorithm Watch төрийн бус байгууллагын үүсгэн байгуулагчийн хувьд 2018 оны Теодер Хойссийн нэрэмжит шагнал). Тэрээр мужийн түвшин дэх хэвлэл мэдээллийн байгууллагууд, сүм хийд, яамдуудад дижиталчлалын нийгэм дэх үр нөлөө гэсэн чиглэлээр зөвлөгөө өгдөг. 2018 оны 9 сараас Бундестагийн „Хиймэл оюун ухаан“ комиссын байнгын гишүүнээр ажиллаж байна.

### Конрад-Аденауэр-Сан

#### Др. Норберт Арнолд

„Боловсрол, шинжлэх ухааны улс төрийн бодлого“ багийн ахлагч

Улс төр, зөвлөгөөний ерөнхий алба

T: +49(0)30 / 26 996-3504

[norbert.arnold@kas.de](mailto:norbert.arnold@kas.de)

Шуудангийн хаяг: Konrad-Adenauer-Stiftung, 10907 Berlin

Эрхлэн нийтлүүлсэн: Конрад-Аденауэр-Сан, 2018 он, Санкт Аугустин, Берлин хот.

Хэвлэлийн зураг төсөл бэлтгэл: yellow too Pasiek Horntrich GbR

Хэвлэлийн эх бэлтгэл: Жанинэ Хөөлэ, Конрад-Аденауэр-Сан

Хэвлэлийн хяналт: Жэнни Калерт, Конрад-Аденауэр-Сан.

ISBN 978-3-95721-496-6

