ХИЙМЭЛ ОЮУН УХААН БА ФИЛОСОФИЙН ХООРОНДЫН МУТУАЛИЗМ

М.Отгонбаяр (Доктор PhD)
С.Бүжинлхам (Магистр)
МУИС, ШУС, Философи, шашин судлалын тэнхим

Түлхүүр үгс: *квалиа, тархи, сэтгэх, оюун санааны философи*

Товч утга: *Тус өгүүлэлд зохиогчид хиймэл оюун ухаан ба философийн салбарт мутуалист буюу харилцан ашигтай харилцаа байна гэж үзэн энэхүү харилцаа нь хэрхэн илэрч байгааг судлахыг зорьсон. Улмаар хиймэл оюун ухаан гэх ойлголтын нэр томъёоны утгыг тодруулж, дараа нь хиймэл оюуны салбарын түүхээс дурдаад оюун санааны философид хамрах машин сэтгэж чадах уу буюу хиймэл агент хүнтэй ижил оюун ухаантай эсэх тухай асуултанд скептик байр сууринаас Ж.Сёрлын "Хятад өрөө"-ний тухай сэтгэхүйн туршилтанд анализ хийсэн болно. Энэхүү анализын үр дүнд хиймэл оюун болон философийн хооронд мутуалист харилцаа нь дараах байдлаар илэрч байна хэмээн дүгнэлээ. Үүнд хиймэл оюун ухааны салбарын онолын нэр томъёог тодруулах, хүрээг татаж өгөх бололцоог философи хангаж байгаа бол философийн уламжлалт асуудлыг шинэ хэлбэрээр асуух боломжийг хиймэл оюун философид олгож байгаа аж.*

* * * * *

Хиймэл оюун ухааны (ХОУ) тухай асуудлууд зөвхөн компьютер ч биш, философи, сэтгэл судлал, нейробиологи, хэл шинжлэлийн асуудалтай нягт уялдаа холбоотой байдаг. Өгүүллийн гарчигт буй мутуализм гэх нэр томъёогоор биологийн ухаанд хоёр өөр зүйлийн амьтан аль аль нь бие биедээ харилцан ашигтайгаар хамтран амьдрах хэлбэрийг нэрлэдэг билээ. Философи болон онолын мэдлэгийн бусад салбарын хооронд мөн ийм "мутуалист" харилцаа байдаг учраас өнөөг хүртэл биологийн, физикийн зэрэг философи оршсоор ирсэн нь тодорхой. Энэхүү өгүүлэлд хиймэл оюун ухаан болон философийн хооронд бусад салбаруудын дунд байдаг шиг тийм мутуалист харилцаа байгаа гэж тооцон, уг харилцаа хэрхэн илэрч байгааг онцгойлон авч үзсэн болно. Философи болон ХОУ-ын хооронд мутуалист харилцаа байгаа гэж үзэх болсон гол шалтгаан нь хиймэл оюун ухааны философи[1] гэх философийн шинэ салбар аль хэдийнээ байр сууриа олсон явдал хийгээд А.Сломаны "ХОУ-ы ихэнхи ажлыг аль хэдийн философичид хийчихсэн байдаг"[2] гэх санаа юм. Өөрөөр хэлбэл философийн үүднээс ХОУ-ы салбарт холбогдох асуудлууд байхын хамт ХОУ-ы түүхийг ч мөн философичдийн бүтээлээс эрэлхийлж болох аж. Энэ утгаараа "Философийн хүрээнд хиймэл оюуныг авч үзэх нь ямар ашиг тусыг хоёр салбарт өгөх вэ?" гэсэн асуулт энэ сэдвийн хүрээн дэх гол асуудал болж байна.

Хэлний философи, оюун санааны философи, шинжлэх ухааны философи, логик, ёс зүй, эпистемологи, метафизик зэрэг салбарын үүднээс хиймэл оюун ухаанд холбогдох асуултуудыг тавьж болох бололцоотой хэдий ч бүгдийг нэг дор асуух гэж оролдох нь нэг өгүүллийн хүрээнд хэтийдсэн ажил болох тул асуудлын хүрээг хиймэл оюун ухааны философид хамаарах асуултуудыг оюун санааны философиор хязгаарлахыг хичээсэн болно.

---

[1] *Matt Carter. Minds and computers: An Introduction to the philosophy of Artificial Intelligence. Edinburg University Press., 2007 тэргүүтэй ном, сурах бичиг хэвлэгдэн гарах болжээ.*

[2] *Aaron Sloman. A Philosophical Encounter. InProceedings14th International Joint Conference on AI Montreal., August 1995.p 124*

Энэ нь "Машин үнэхээр сэтгэж чадах уу?", "Хүний болон машины оюун ухаан яг адил уу? "Хүний тархи компьютер гэсэн үг үү?" "Машинд мэдрэмж байх бололцоотой юу?" гэх зэрэг асуултуудыг анхаарна гэсэн үг юм.

Нэр томъёоны тодруулга:

Хиймэл оюун ухааны түүхийг дурдахаас урд уг ойлголтыг ямар утгаар хэрэглэдэг болохыг тодруулах нь зүйтэй билээ. "Хиймэл оюун ухаан бол компьютерт оюун ухаантай гэж тооцогдох авир үйлдлийг үзүүлэх боломжийг олгогч программ хөгжүүлэхийг зорьдог компьютерийн шинжлэх ухааны салбар юм"[3] гэж Станфордын философийн нэвтэрхий тольд тодорхойлсон байдаг бол Жон Маккарти "Ухаалаг машин тэр дундаа ухаалаг компьютерийн программыг бүтээхийг зорьдог инженерчлэлийн болон шинжлэх ухааны мэдлэг юм"[4] гэж тодорхойлжээ. Стюарт Рассель болон Петер Норвиг нар хиймэл оюун ухаанд

1. Хүнтэй адил сэтгэдэг
2. Хүн шиг үйлддэг
3. Рационалиар сэтгэдэг
4. Рационалиар үйлддэг[5] гэж авч үзэх дөрвөн янзын хандлага байна хэмээн бүлэглэн авч үзсэн байна.

Түүнчлэн шатар тоглох, оношилгоо хийх, тоо бодох зэрэг тодорхой нэг төрлийн үйлдлийг л хийх чадвартай программ бүтээхийг сул хиймэл оюун ухаан (weak artificial intelligence), хүнтэй адил ухаалаг машины тухайд ерөнхий хиймэл оюун ухаан(general artificial intelligence)[6] гэж нэрлэдэг ажээ. Цаашилбал Ник Бостром хиймэл супер оюун ухаан (Artificial Super Intelligence) гэдэг ойлголтоор хүнээс хавьгүй илүү ухаалаг байх бололцоот оюун ухааныг илэрхийлэх болсон байна. Гэтэл энд нэг асуудал байгаа нь ерийн хэрэглээнд хиймэл оюун ухаан гэдэг үгийг ухаалаг машиныг бүтээхийг зорьдог салбар гэхээсээ илүүтэй уг салбарын бүтээл буюу оюун ухааныг өөртөө агуулж буй машин эсвэл программ гэх утгаар ойлгодог билээ. Жишээлбэл, "TED talk"-ийн илтгэгч, блог хөтлөгч Тим Урбан л гэхэд "Робот бол хиймэл оюун ухааныг тээгч юм... Хиймэл оюун ухаан гэдэг бол өөрөө роботын дотор буй компьютер нь юм"[7] хэмээн "Хиймэл оюун ухааны хувьсал, Супер оюун ухаанд хүрэх зам*"* гэх нийтлэлдээ бичжээ.

Нөгөө талд хиймэл оюун ухаан судлаачид "үйлдэл хийж буй аливаа зүйлийг агент (agent)"[8] хэмээн үздэг аж. Улмаар гадаад орчноо хүлээн авч хариу үйлдэл үзүүлж буй тохиолдолд ухаалаг агент (intelligent agent), хамгийн сайн үр дүнд хүрэхийг зорин үйлдэл хийдэг бол рациональ агент гэх зэргээр нэрлэдэг байна. С.Рассель, П.Норвиг нарын бүлэглэсэн дөрвөн хандлагын алийг баримталж буйгаас шалтгаалан хиймэл оюун ухаан судлаачдын бүтээсэн үр дүн болох машин, программ, системийг ухаалаг агент, рациональ агент, хиймэл агент, хиймэл оюуны систем, абстракт ухаалаг агент гэж өөр өөрөөр нэрлэх бололцоотой байдаг ажээ Ийнхүү нэр томъёоны хувьд төөрөгдөл үүсэхээс сэргийлэн цаашид Ж.Лайрд болон П.Росенблум нарын судлаачдын хэрэглэсэн "хиймэл агент"[9] гэх ойлголтоор ХОУ-ы салбарын үр дүнг нэрлэе.

---

[3] *Richmond Thomason. "Logic and Artificial Intelligence". The Stanford Encyclopedia of Philosophy. Winter 2016 Edition. Edward N. Zalta (ed.), URL=* https://plato.stanford.edu/entries/logic-ai/index.html#jmc

[4] *John McCarthy. What is Artificial Intelligence. Revised November 12, 2007. URL=* http://www-formal.stanford.edu/jmc/whatisai/

[5] *Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (Third edition). Pearson., 2010. p 1*

[6] *Дэлгэрүүлж Ben Goertzel, Cassio Pennachin. Contemporary Approaches to Artificial General Intelligence (pp 1-30) гэх өгүүллийг Goertzel, Ben, Pennachin (Eds). Artificial General Intelligence. Springer., 2007 номноос үзнэ үү.*

[7] *Tim Urban. The AI Revolution: The Road to Superintelligence. January 22, 2015. URL=* http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

[8] *Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (Third edition). Pearson., 2010. p 4*

[9] *J.E. Laird and P.S.Rosenbloom. Report on the AAAI 1991 Spring Symposium on "Integrated Intelligent Architectures". AI Magazine:12. 1991.p 35*

Хиймэл оюун ухааны түүхээс:

Хиймэл оюун ухааны түүхийн хувьд 1956 онд Жон Маккарти, Марвин Минский тэргүүтэй компьютер, когнитив шинжлэх ухааны судлаачдын Дартмут коллежид зохион байгуулсан эрдэм шинжилгээний хурал дээр уг нэр томъёог анх хэрэглэж, энэ салбарын зорилго, зорилтыг тодорхойлсон байдаг[10]. Гэхдээ хиймэл агент гэгч зүйл байж болох тухай санааг хүн төрөлхтөн аль эрт бодож олсон байдаг. Тухайлбал эртний Грекийн Гефест бурханы тухай домогт эзэндээ үйлчлэгч алтан роботуудын тухай гардаг бол XIX зуунд Мари Шэллей "хиймэл хүн" болох Франкенштейны тухай зөгнөлт зохиол бичиж, 1950 онд Исаак Азимов роботын тухай гурван хуулийг өөрийн бүтээлдээ дурджээ. Хэдийгээр утга уран зохиолд хиймэл хүн, ухаалаг агент байж болох юм гэсэн санааг нийтэд түгээж байсан ч яаж хэрхэн эдгээрийг бий болгох вэ? гэсэн асуултанд бүрэн дүүрэн хариулт өгөөгүй байдаг.

Харин ирээдүйд ХОУ гэж нэрлэгдэх уг салбарын суурь онолын үр хөврөлүүдийг философичид, байгаль шинжээчид өөрсдийн зохиол бүтээлдээ илэрхийлж байжээ. Жишээлбэл Эртний Грекийн сэтгэгчид болох Парменид, Протагор, Сократ, Платон, Зенон, Хриспи нарын үзэл хийгээд Аристотелийн зохиол бүтээл нь одоогийн компьютерийн шинжлэх ухаан үндэс сууриа болгож буй логикийн шинжлэх ухааныг (ерөнхий суурийг) хүн төрөлхтөнд "бэлэглэсэн" бол Р.Декарт XVII зуунд өөрийн дуализмаараа бие оюун санааны хоорондын харилцааны тухай асуудлыг сэргээж, түүнтэй хамт Т.Гоббс, Г.Лейбниц нар рациональ бодол санаа нь алгебр, геометртэй адил системлэг байж болох талаар судалж байв. Хүн төрөлхтөн механикийн хуулиудыг нээж хүн ч, амьтан ч өөрөө машин юм гэх үзэл хүртэл гарч ирсэн. Улмаар Б.Паскаль анхны механик дижитал тооцоолох машиныг бүтээж, Г.Лейбниц Б.Паскалийн машинд үржих хуваах үйлдлийг нэмэн сайжруулж, Б.Рассель, А.Н.Уайтхед нарын хэвлүүлсэн "Математикийн зарчмууд" бүтээл нь формал логикт хувьсгал хийж, А.Тюринг сэтгэдэг машин бүтээх бололцооны талаар судалсан зэргээр үе үеийн эрдэмтэд өмнөх үеийхнийхээ санааг уламжлан авч хөгжүүлсээр орчин үеийн компьютер, робот, программ хангамж зэргийг бүтээсэн байна.

Анхны үйлдлийн компьютерийг А.Тюрингийн багийн 1940 онд бүтээсэн Хеат Робинсон хэмээх машин гэж үздэг байна. Үүний дараа жил Германд зохион бүтээгч Конрад Зус Z-3 нэртэй програмчилж болохуйц компьютерийг бүтээжээ.[11] 1955-1956 онд А.Ньюэлл, Г.А.Саймон, К.Шоу нарын бүтээссэн учир зүйг тодорхойлдог "Logic Threorist" хэмээх программыг анхны сэтгэж чадах программ гэж үздэг байна.[12] 1956-1970-д оны хооронд судлаачид ХОУ-ы талаар өөдрөг үзлүүдийг дэвшүүлж байсан агаад жишээлбэл Г.А.Саймон "20 жилийн дотор машин хүний хийж чаддаг бүхнийг чадах болно."[13] гэж 1965 онд тунхаглаж байжээ. Гэвч энэхүү оптимист байр суурь нь маш их хүлээлтийг бий болгохын зэрэгцээ, нүүр царай ялгах, орон зай баримжаалах зэрэг нь хиймэл агентын хувьд хүндрэлтэй үйлдлүүд болохыг ажиглаж эхэлсэн байна. Өөрөөр хэлбэл машины хязгаарыг ойлгож эхэлсэн байна. Тэгээд 1980-д оноос хойш ХОУ нь аж үйлдвэржил гэдэг утгаараа хөгжин шинжлэх ухааны аргазүйг ашиглах болж, 2001 оноос эхлэн их хэмжээний өгөгдлийн (data) олонлогийг ашиглах бололцоотой болсон байна. Улмаар өдгөө тоглоом тоглодог, тодорхой хүрээнд асуудлыг шийдвэрлэдэг, алгебрын бодлого бодож, өвчин оношлох зэрэг хүний хийх бололцоотойгоос эхлээд бололцоогүй олон төрлийн чадварыг эзэмшсэн хиймэл агентууд бүтээгдсээр байна. Жишээлбэл 2015 онд л гэхэд "Хансон Роботикс" компани нүүрний 62 хувирал илэрхийлэх, харилцан яриа өрнүүлэх чадвартай Софиа хэмээх хүн дүрст роботыг бүтээсэн агаад түүнд 2017 оны 10 сард Саудын Араб иргэншил олгосон. Түүгээр үл барам тухайн оны 12 сард BINA48[14] гэх Софиатай ижил үйлдвэрийн робот коллежийн "Хайрын

[10] *Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (Third edition). Pearson., 2010. p 2*

[11] *Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (Third edition). Pearson., 2010. p 14*

[12] *Daniel Crevier. AI: The Tumultuous Search for Artificial Intelligence. New York., 1993. p. 44*

[13] *Daniel Crevier. AI: The Tumultuous Search for Artificial Intelligence. New York., 1993. p. 109*

[14] *Articles about BINA48 url;* http://www.hansonrobotics.com/category/news/bina-48/

философи" гэсэн хичээлийн шалгалтанд тэнцсэн билээ.

Хиймэл оюун ухаанд яагаад философи хэрэгтэй вэ?

ХОУ-д холбогдолтой "Машин сэтгэж чадах уу?", эсвэл "Машин Х үйлдлийг хийж чадах уу? гэх асуудлыг өдгөө хиймэл оюун ухаанд холбогдох сонгодог онолын (classical theoretical)[15] асуудал хэмээн үзэх болсон байна. Энэ удаад сонгодог онолын хүрээнд багтах "Машинд квалиа бий юу?" "Хүний болон машины оюун ухаан адил уу?" "Тархи нь дижитал компьютер юм уу?" зэрэг философид хамааралтай асуултууд нь хиймэл оюун ухаан судлаачид, нейробиологичид, болон буцаад философид өөрт нь ямар ач холбогдолтойг тодруулан гаргахын тулд ерөнхийдөө хүнтэй яг адил хиймэл агент байх бололцоогүй гэж үзэн, машины "сэтгэдэг" эсэхэд эргэлзсэн няцаалтын ач холбогдлыг авч үзье.

Хятад өрөө буюу Ж. Сёрлын шүүмж

Жон Сёрл ХОУ-г сул, хүчирхэг *(Уг ангилал нь сул болон ерөнхий хиймэл ухаан гэдэг ойлголтуудтай зарим талаараа давхцдаг байна.)* хэмээн хуваасан байна. ХОУ нь бидний хувьд зөвхөн хэрэгсэл байх агенттай хамаатай бол сул ХОУ, яг хүний түвшинд сэтгэж, хүнтэй адил оюун санааг эзэмших эсэхтэй холбоотой асуудлыг анхаардаг бол хүчирхэг ХОУ гэж үзжээ. Улмаар Ж.Сёрл хүчирхэг ХОУ байх боломжгүй юм гэсэн санаа дэвшүүлсэн бөгөөд уг үндэслэлээ дэмжихийн тулд хүний асуултанд хариулах чадвар бүхий Рожер Шанкийн загвараар *(1969 онд Р.Шанк төрөлх хэлийг ойлгоход зориулсан ойлголтоос хамаарсан загварыг /conceptual dependency model/ танилцуулсан байна.)* жишээ татаж улмаар хятад өрөөний тухай сэтгэхүйн туршилтаа дэвшүүлсэн байдаг. Дашрамд дурдахад уг "сэтгэхүйн туршилт" нь Р.Шанкийн бүтээл төдийгүй хүний оюуны үзэгдлийг дуурайсан А.Тюрингийн тестийг[16] давах чадвар бүхий аливаа машины эсрэг чиглэсэн байдаг юм.

Ж.Сёрл хятад өрөөг хэрхэн дүрсэлснийг товчоор өгүүлвэл: Тэрээр нэг өрөөнд хятад ханз бичээс бүхий гурван багц цаас, эдгээрийг хэрхэн холбох тухай заавар бүхий номтой (мэдээж англи хэл дээрх) түгжигджээ. Тэр хятад хэл огтоос гадарлахгүй агаад түүний хувьд эдгээр бичээс нь зүгээр л дүрсүүд юм. Түүний хийх ёстой зүйл гэвэл зааврын дагуу гурван багцийг холбох явдал юм. Багц бичээсүүдийн нэг нь түүх, нөгөө нь асуулт, бас нэг нь хариулт байгаа ч Ж.Сёрл хятад хэл мэдэхгүй тул зүгээр л зааврын номын дагуу дүрсээр нь харж "эвлүүлнэ." Өрөөний гадна байгаа хэн нэгэн хятад хэл мэддэг хүн түүний холбож нийлүүлсэн бичээс бүхий цаасыг (асуулт хариултыг) уншвал Ж.Сёрл хятад хэлийг эзэмшсэн нэгэн мэт харагдана. Гэтэл үнэндээ тэр хятад хэлийг огт мэдэхгүй, тухайн бичээсийн утгыг ойлгохгүй зөвхөн зааврыг л дагаж байгаа хэрэг юм.

Компьютерийн программын ажиллах зарчим нь хятад өрөөнд байгаа хүнтэй адил юм. Учир нь Р.Шанкийн загварын дагуу бүтээсэн программ аливаа бичээсийг уншаад, холбоо харилцааг "ойлгоно" гэдэг нь хүн төрөлх хэлээрээ бичээсийг уншаад ухаарч ойлгож байгаатай адилгүй. Иймээс компьютер асуудлыг шийдэх, "эх хэлийг ойлгох" зэрэг нь оюун ухаантай байгаа мэт л дүр үзүүлж буй хэрэг гэсэн үг хэмээн Ж.Сёрл үзсэн байна. Улмаар Ж.Сёрл асуудлыг илүү тодорхой болгохын тулд дараах тайлбарыг өгсөн байдаг. Түүний хувьд герман хэлийг ойлгох түвшин нь англи хэлийг ойлгох түвшнээс арай доогуур байгаа ч ямар нэг хэмжээгээр ойлгож байгаа бол компьютерийн хувьд энэ чадвар тэг байдаг гэжээ. Учир нь Ж.Сёрлын үзлээр аливааг ойлгох нь шалтгаант холбоог мэдэх, ялгаж таних чадвартай холбоотой агаад тархины шалтгаант шинжийн (feature) үр дүн бол зорилгот төлөв (intentionality) ажээ. Эндээс аливааг ойлгох чадвартай байхын тулд зорилгот төлөвтэй байх ёстой гэсэн санаа гарч байна. Зорилгот төлвийг бий болгох чадвартай аливаа механизм нь тархины шалтгаант холбоог ухаж ойлгох чадвартай ижил механизмтай байх

[15] *Vincent C. Müller (Ed.) Philosophy and Theory of Artificial Intelligence. Springer., 2013. p vii*

[16] *А.Туринги йн тестийг хиймэл агентийг ухаалаг авир үйлдэл гаргаж чадаж буй эсэхийг шалгах шалгуур гэж товчоор тодорхойлж болно. Дэлгэрэнгүйг A. M. Turing. Computing Machinery and Intelligence. Mind 49: (pp 433-460). 1950 өгүүллээс харна уу.*

ёстой. Гэтэл формал тэмдэгтүүдийн дагуу ажилладаг программын хувьд уг чадвар байх бололцоогүй. Учир нь цэвэр формал загвар нь зорилгот төлвийн хүрэлцээтэй нөхцөл биш агаад, дангаараа дараагийн шатны формализмын шалтгаан нь болж чадахгүй. Программын хийж буй аливаа үйлдэл бол тэмдэгтүүдийг чиглүүлэх (manipulate) үйлдэл төдий агаад программын хувьд тэмдэгтүүд нь тэмдэглэгдэгчгүй тул утга илэрхийлэхгүй. Өөрөөр хэлбэл семантик үгүй синтактикийн бүрдэл тул зорилгот төлөв хийгээд ойлгохуйн тухай асуудлыг программд хамаатуулах аргагүй юм. Программд синтактикийн бүхэл бүтэн систем бий агаад программ хүний оюуны үзэгдлийг загварчилдаг. Гэхдээ синтактик дангаараа семантикийг бүрдүүлэхэд хүрэлцээгүй агаад, загварчлал гэдэг бол хуулбарлан буулгаж байгаа явдал биш юм. Товчхондоо программ (хиймэл агент) нь бүхэлдээ зорилгот төлөв бүхий субьектийн хэрэгсэл болохоос бус зорилгот төлвийг , шалтгаалцын эх үүсврийг бүтээгч огтоос биш гэж хэлж болох юм.

Дээр дурдсан Софиа болон BINA48 зэрэг хүн дүрст роботуудын тухайд ч Ж.Сёрл тэд төрөлх хэлээр харилцаж байгаа нь тэднийг хүнтэй ижил түвшний оюун ухаантай, ойлгохуйтай гэсэн үг биш гэж хэлэх байсан биз. Гэтэл энд нэг бяцхан асуудал байгаа нь BINA48 робот энгийн нэг хиймэл агент биш "ухамсар" бүхий робот билээ. BINA48 роботод түүний эзэмшигч Мартин Ротблаттын эхнэр Бина Ротблаттын ой дурсамж, мэдрэмж, итгэл үнэмшил зэргийг "оюун санааны файл" (mindfile) хэлбэрт шилжүүлэн суулгасан билээ. Өөрөөр хэлбэл энэхүү робот нь хэн нэгний тухай нарийвчилсан мэдээллийг цуглуулж, түүнийгээ оюун санааны файл гэх дижитал хэлбэрт оруулах замаар хүний ухамсрыг бүтээж түүнийгээ биологийн эсвэл технологийн замаар бий болгосон биетэд шилжүүлэх боломжтой эсэхийг туршсан туршилтын загвар юм. Тэгвэл энэ BINA48-д Ж.Сёрлын хүсээд байсан зорилгот төлөв, семантик бүхий синтактик утгыг ойлгох чадвар байх боломжтой гэж үзэх үү? гэдэг асуулт урган гарна. Учир нь BINA48 дуурайх үйлдлийг хийх төдийгүй түүнд хүнээс "хуулбарлан буулгахыг оролдсон" оюун санааны файл байна. Дээрх асуултын хариултыг ухамсрыг тодорхойлсон Ж.Сёрлын тайлбараас эрье. Ж.Сёрл оюун санаа, ухамсрын тухайд өгсөн тайлбарыг биологийн натурализм гэж нэрлэдэг аж.

Ж.Сёрл "Ухамсар" гэх өгүүлэлдээ "Минийхээр ухамсар гэдэг бол бидний өглөө сэрээд орой унтах хүртэл байх сэрэхүйн буюу мэдээлэлтэй байх төлвүүд (states of sentience or awareness) юм"[17] гэж бичжээ. Тэгэхээр ухамсар гэдэг бол статик мэдээ төдий биш харин динамик үйл явц аж. Түүний үзснээр ухамсар нь чанарын, субьектив, нэгдмэл онцлог шинжээс бүрддэг байна. Ухамсрын төлөв бүрт тодорхой мэдрэгдэх чанар байдаг. Бялууг амтлах болон шувууны жиргээг сонсох нь тэс өөр мэдрэмжийг өгнө гэдгийг хүн бүр мэднэ. Мэдрэхүйн туршлага төдийгүй сэтгэхүйн үйл явц ч үүнд хамаарна. Жишээ нь монголоор "Тэр бол бүсгүй хүн" гэж хэлэх, францаар "Elle est la femme" гэж хэлэх ч тэс өөр санагдах билээ. Энэхүү дотоодод ялгаатай мэдрэгдэх байдлыг Ж.Сёрл чанарын онцлог шинж гэсэн бол Р.Нажел тэргүүтэй философичид квалиа гэж нэрлэсэн байдаг. Квалиагийн тухай товчхон дурдахад уг ойлголтыг ухамсрын тухай хэцүү асуудал (hard problem of consiousness) хэмээн нэрлэдэг агаад Кларенс Ирвинг Левис 1929 онд анх квалиа гэдэг ойлголтыг хэрэглэсэн байна. Орчин үед

1. Квалиа бол үзэгдэлт чанар
2. Квалиа бол сэрлийн мэдээллийн шинж
3. Квалиа бол дотоод, репрезентаци хийх боломжгүй шинж
4. Квалиа бол дотоод, физик бус, үгээр илэрхийлэх боломжгүй шинж[18] гэсэн дөрвөн утгаар квалиа гэх нэр томъёог хэрэглэх болсон.

Ж.Сёрлийн авч үзсэн чанарын онцлог шинжийг гурав дахь бүлэгт багтаах боломжтой

[17] *John Searle. Consciousness. url;* http://faculty.wcas.northwestern.edu/~paller/dialogue/csc1.pdf, *p3*

[18] *Tye, Michael, "Qualia", The Stanford Encyclopedia of Philosophy url;* https://plato.stanford.edu/archives/win2017/entries/qualia/

юм. Гэхдээ тэрээр квалиа нь ухамсрын төлөв тул тусад нь ялгаж нэрлэх нь зохимжгүй гэж үзсэн агаад чанарын онцлогт сэтгэхүйн төлвийг нэмж багтааснаараа онцлог байгаа юм.

Энэхүү чанар гэх онцлог шинжээс субъектив байдал хамааралтай байдаг байна. Тайлбарлаваас яг энэ мөчид та алим хазлаа гэхэд тухайн "онцгой" амтыг та л мэдрэхээс бус өөр хэн нэгэн мэдрэх бололцоогүй юм. Энэ утгаараа ухамсар бол нэгдүгээр биеийн онтологи *(оршихуйн горим гэдэг утгаар Ж.Сёрл онтологи гэдэг үгийг ашигласан)* бөгөөд "ямар нэгэн тархи" л мэдэрч байж ухамсрын төлөв оршдог байна. Нэгдмэл шинж гэдэг нь тусгаар тусгаар сэрлүүд хамтдаа нэгдсэн ухамсрын талбарын нэг хэсэг болж байгааг онцолж буй хэрэг юм. Жишээлбэл та компьютер дээр текст бичлээ гэхэд танд компьютерийн дэлгэц дээрх хар цагаан дүрс, компьютерийн хатуу товчлуур, уг товчийг тогшиход гарах дуу зэрэг нь тус тусдаа бус бүхлээрээ мэдрэгдэнэ шүү дээ.

Цаашилбал, тархины нейроны доод түвшинд болж буй процессоос ухамсрын процесс шалтгаалдаг, харин ухамсар нь тархины бүтцийн дээд түвшний процессоор тайлбарлагдана гэж Ж.Сёрл үзжээ. Түүнийхээр оюуны үзэгдэл бол тархины онцлог шинж агаад өмнө дурдсан зорилгот төлвийг бий болгох чадвар тархинд л байдаг байна. Бүр цаашилбал аливаа оюуны үзэгдэл нь зарим талаараа митоз, мейоз хуваагдал, хоол боловсруулалт зэрэгтэй адил биологи үйл явц байдаг аж. Сонирхолтой нь Ж.Сёрл хүнийг машин гэж үзсэн байна. Ж.Сёрлын энэхүү үндэслэл нь Р.Декартын бие оюун санааны тухай дуализмыг санагдуулах авч тэрээр оюун санаа бол объектив байдлаар салган авч үзэж болох эд биш гэж үзэн дуализмыг эсэргүүцжээ. Учир нь ухамсар нэгдүгээр биеийн онтологи тул үүнийг гуравдугаар биеийн өнцгөөс харахын тулд таналт хийх бололцоогүй юм. Өөрөөр хэлбэл түүний хувьд оюун санаа гэдэг бол машинд суулгасан сүнс/программ биш харин тухайн машины өөрийнх нь үйл ажиллагааны үр дүн ажээ. Ухамсрыг бий болгож буй тархины микро түвшний үйл ажилллагааг судлаачид судалж болох ч өвдөлт, хүсэл, итгэл үнэмшил зэрэг субьектив төлвийг томъёолж тооцоолж чадахгүй билээ.

Тэгвэл хүн төрөлхтөн тархийг тэрхүү онцлог шинж, чадвар чадамжийнх нь хамт хуулбарлах юм бол "хиймэл машин" сэтгэж чадна гэсэн үг үү? гэж асууж болох юм. Энэхүү асуултын тухайд Ж.Сёрл 2016 онд тавьсан "Хиймэл оюун дахь ухамсар"[19] гэх илтгэлдээ энэ бол нээлттэй асуудал хэмээн хариулсан байдаг. Гэхдээ одоогийн байдлаар тархийг бүрэн судалж, нууцыг нь нээгээгүй байгаа тул биологийн натурализмын үүднээс хиймэл агент сэтгэж чадна гэж хариулах бололцоогүй гэдэг нь тодорхой байна.. Дээр дурдсан BINA48-ын хувьд түүнд хэдий Бина Ротблатын дурсамжаас бүрдсэн оюун санааны файл байгаа хэдий ч түүнийг ухамсартай гэж үзэх боломжгүй юм. Учир нь ухамсар бол динамик төлөвт байдаг үйл явц агаад, түүнийг объектив байдлаар судлах нь учир дутагдалтай болохыг ухамсрын субъектив шинжийн тухай тайлбар гэрчилж байна. Тэгэхээр BINA48-д зорилгот төлөв байхгүй тул түүний асуултанд хариулна гэдэг нь тэмдэгтүүдийг арай "боловсронгуй замаар" чиглүүлж байгаа явдал л болж таарч байна.

Дээрх тайлбаруудаас дүгнэхэд программистууд хичнээн сайн программ зохиолоо ч тэр нь өөрийгөө ойлгож байгаа гэдгийг Ж. Сёрлд ч, бусдад ч бүрэн дүүрэн хангалттай хэмжээнд итгүүлж чадахгүй учраас хятад өрөөний асуудал нь болон ухамсрын тухай Ж.Сёрлын тайлбар нь хиймэл оюун ухаан бүтээхийг зоригчдын хувьд энэ тэргүүний чухал ач холбогдолтой үндэслэл биш байж болно. Учир нь одоогоор хүн төрөлхтөн бусдын санаа бодлыг бүрэн дүүрэн унших арга замыг хараахан олоогүй л байна. Хүмүүст бага ч атугай ойлгосон гэдгээ илэрхийлэхийн тулд бид үг яриагаа ашигладаг ч компьютерийн "гаралт, илэрхийлэл"-ийг Ж.Сёрл бүрэн дүүрэн ойлгосны үр дүн биш гэж үзсэн шүү дээ. Тиймээс "компьютерийн санаа бодлыг унших" төхөөрөмж бүтээж байж л хиймэл оюун ухаан нь сэтгэж байгаа мэт дүр үзүүлээгүй, харин жинхэнээсээ сэтгэж байгаа гэдэгт итгэлтэй болно.

---

[19] *John Searle. Consciousness in Artificial Intelligence. url;* https://www.youtube.com/watch?v=rHKwIYsPXLg

ХОУ судлаачид уг асуултад хариулахын тулд хүлээхгүйгээр урагшилсаар байгаагийн жишээ бол дээр өгүүлсэн Софиа болон BINA48 гэх роботууд билээ. BINA48-ыг бүтээсэн явдал нь хүний ухамсрыг дижитал хэлбэрт оруулах бололцоотой юу? Оюун санаа гэдэг ой дурсамжийн хэлхээ юу? Субьектив туршлагыг хуулбарлах бололцоотой юу? Ухамсар болон оюун санааг хэрхэн ойлгож тодорхойлбол зохих вэ? гэдэг асуултыг араасаа дагуулж байгаа нь харагдлаа. Түүнчлэн дээр дурдсан хиймэл агент ойлгож байгаа гэдгээ итгүүлэх тухай уг асуудал нь ондоо хүний оюун санааны тухай асуудалтай холбогдож болно. Үнэн хэрэгтээ хүмүүс өөрөөсөө бусдыг надтай адил оюун санаатай гэдэгт интерспекци хийж л итгэж байгаа билээ. Хиймэл агентын хувьд уг асуудал нь түүнийг хүн өөртэйгээ адилтган улмаар интроспекцийн үндэслэлээ хиймэл агентад "тулгах" ямар боломж, бололцоо байгаа вэ? гэх асуулт урган гарч ирэх билээ. Тэгэхээр хиймэл оюуны салбар дахь шинэ бүтээл философийн уламжлалт асуултыг дагуулж байна.

Мөн хиймэл оюуныхны асуудаг "Машин сэтгэж чадах уу?" гэх асуултанд философичдийн өгөх хариулт нь хиймэл оюун судлаачдыг шинэ сорилтонд дуудаж байгаа нь Ж.Сёрлын тархийг хуулбарлах бололцооны тухай эргэцүүллээс харагдлаа.

Дүгнэлт: Хэдийгээр хиймэл оюун ухааныг философийн үүднээс авч үзсэн бүхий л асуудлыг нарийвчлан авч үзэж бүрэн дүүрэн илэрхийлээгүй ч хүнтэй ижил түвшинд сэтгэдэг хиймэл агент байж болох эсэхэд эргэлзээтэй хандсан "Хятад өрөө" хэмээх сэтгэхүйн туршилтаас харахад хиймэл оюуны салбар болон философийн хооронд мутуализм байна гэх үндэслэл гарч ирж байна. Хиймэл оюун ухаанд философи нь дараах зүйлсийг тодруулахад тус дөхөм болно.

Хиймэл оюун ухааны судалгааны зорилго: Жишээлбэл оюун ухаантай машин гэж юуг хэлэх вэ? Хиймэл оюун ухаан гэж юу юм бэ? гэх мэт энэ салбарт ашиглах ойлголтуудын тодорхойлолтыг философиос эрэх боломжтой байна. Өөрөөр хэлбэл философи нь хиймэл оюун ухааны салбарын онолын нэгэн эх сурвалж болж байна.

Хиймэл оюун ухааныг илүү хөгжих, дэвших, ахихад шаардлагатай хиймэл агентыг шаардахын зэрэгцээ нөгөө талаас зарим утгаараа хиймэл оюун ухааны судалгаанд дохио өгч, хязгаар хүрээг татаж байна.

Философид хиймэл оюун ухаан нь:

Хуучин философийн асуудлуудыг шинэ хэлбэрт оруулан тодорхойлохыг шаардаж байна. Жишээ нь ондоо хүний оюун санааны тухай асуудал хиймэл агентад үйлчлэх үү?

Хүний ухамсар, танин мэдэхүйн тухай асуудлуудыг шинэ шатанд нарийвчлан судлах боломж бололцоог олгож байна.

Судлах шинэ үзэгдлүүд, философийн хуучин үзэл онолуудыг шүүн тунгаах шаардлагыг гаргаж ирэх зэрэг "шинэ шинэ" ажил санаагаар философийг баяжуулж байна гэх мэт олон талаар ач холбогдолтой байна.

ABSTRACT

Artificial intelligence is an interdisciplinary field which intercross with philosophy, ethics, neuro science and computer science. In this article, philosophical approach of artificial intelligence has been discussed. New branch philosophy of artificial intelligence has emerged from scientists and philosophers many years' effort. Can machine think? is a classic question of artificial intelligence. Regarding to this question, some philosophers give positive answer and others negative. For example, John Searle has doubts strong artificial intelligence's existence. In this article, John Searle's argument against artificial intelligence has been analyzed in order to find mutual relation between artificial intelligence and philosophy. Studying artificial intelligence from the point of philosophy rises new problems for philosophy and gives methodical foundation to artificial intelligence.

НОМ ЗҮЙ

1. Философи шашин судлал сэтгүүл XI №336/60/. УБ., 2010
2. Aaron Sloman. A Philosophical Encounter. In Proceedings 14th International Joint Conference on AI Montreal., August 1995
3. Alan M. Turing. Computing Machinery and Intelligence. Mind 49: (pp 433-460). 1950
4. Daniel Crevier. AI: The Tumultuous Search for Artificial Intelligence. New York., 1993
5. Goertzel, Ben, Pennachin (Eds). Artificial General Intelligence. Springer., 2007
6. J.E. Laird and P.S. Rosenbloom. Report on the AAAI 1991 Spring Symposium on "Integrated Intelligent Architectures". AI Magazine:12. 1991
7. John McCarthy. What is Artificial Intelligence. Revised November 12, 2007. URL= http://www-formal.stanford.edu/jmc/whatisai/
8. John Searle. (1980) Minds, brains, and programs. Behavioral and Brain Sciences 3 (3): 417-457
9. John Searle. Consciousness. URL= http://faculty.wcas.northwestern.edu/~paller/dialogue/csc1.pdf,p3
10. John Searle. Consciousness in Artificial Intelligence. URL= https://www.youtube.com/watch?v=rHKwIYsPXLg
11. Matt Carter. Minds and computers: An Introduction to the philosophy of Artificial Intelligence. Edinburg University Press., 2007
12. Richmond Thomason. "Logic and Artificial Intelligence". The Stanford Encyclopedia of Philosophy. Winter 2016 Edition. Edward N. Zalta (ed.), URL= https://plato.stanford.edu/entries/logic-ai/index.html#jmc
13. Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (Third edition). Pearson., 2010
14. Tim Urban. The AI Revolution: The Road to Superintelligence. URL= http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html
15. Tye, Michael, "Qualia", The Stanford Encyclopedia of Philosophy URL= https://plato.stanford.edu/archives/win2017/entries/qualia/
16. Vincent C. Mьller (Ed.) Philosophy and Theory of Artificial Intelligence. Springer., 2013