

Bírálat

Héder Mihály: *Az MI etika új hullámai*
MTA doktori értekezés (2024)

Héder Mihály doktori értekezése időszerű, ambiciózus és koncepcionálisan átgondolt munka, amely a mesterséges intelligencia etikájának jelenlegi diskurzusát nem pusztán leírni, hanem történeti és filozófiai kontextusba ágyazva kritikailag elemezni kívánja. A dolgozat terjedelme és tematikai gazdagsága önmagában is jelzi, hogy szerzője nem részproblémák megoldására vállalkozott, hanem az MI-vel kapcsolatos normatív reflexió egészének szerkezetét kívánja feltérképezni és újrarendezni.

A munka egyik fő érdeme, hogy nem elégszik meg a jelenlegi etikai irányzatok tipologikus felsorolásával, hanem azokat történeti, technikatörténeti és műszaki-episztemológiai keretbe illeszti. A Bevezetésben megfogalmazott „múlt-sokk” motívum – az a felismerés, hogy a ma újnak tűnő dilemmák jelentős része már évtizedekkel korábban is megjelent – különösen termékeny perspektívát kínál. Ez a történeti érzékenység végigkíséri az értekezést, és megóvja a szerzőt attól a naiv jelenorientáltságtól, amely az MI-vel kapcsolatos közbeszédet gyakran jellemzi.

A dolgozat fő erényei

1. A konstruktóri perspektíva

A dolgozat egyik legfontosabb és legeredetibb hozzájárulása az úgynevezett „konstruktóri szemszög” következetes alkalmazása (1.1. fejezet). A szerző világossá teszi, hogy az MI etikai problémái nem érthetők meg a fejlesztési folyamat sajátosságainak figyelembevétel nélkül. A dekompozíció kényszere, a moduláris felépítés, a csapatmunka és az episztemikus korlátok együttese olyan sajátos helyzetet teremt, amelyben a normatív döntések jelentős része a „tervezési időben” születik meg.

Különösen erős a felelősségrés elemzése az önvezető járművek kapcsán (1.2. fejezet). A szerző meggyőzően mutatja be, hogy a sofőr eltűnése nem egyszerűen jogi vagy morálfilozófiai problémát generál, hanem az episztemikus távolság radikális megnövekedését is jelenti. A konstruktóri csapat helyzete időben, térben és szituációs tudás tekintetében is eltér a hagyományos felelősségi struktúráktól. Az „episztemorális” tervezési döntések fogalma ezzel összefüggésben kifejezetten termékeny újítás.

2. A műszaki tudás filozófiai elemzése

A 2. fejezet a műszaki tudás természetének elemzésével megalapozza a későbbi etikai fejtegetéseket. A technológia normatív, célszerű és prediktív jellegének hangsúlyozása fontos ellensúlyt képez a tisztán leíró tudományideálokkal szemben. A szerző helyesen mutat rá arra, hogy minden mérnöki tevékenység implicit jövőtervezés, s így inherens morális dimenzióval rendelkezik.

A tervezés mint predikció gondolata különösen fontos: a mérnöki terv mindig előrejelzés arra vonatkozóan, hogy egy bizonyos struktúra működni fog. Ezzel a szerző világosan elhatárolja a műszaki tudást a pusztán deskriptív tudásformáktól.

3. A történeti és kanonizációs reflexió

A 6–9. fejezetekben a szerző az MI-etika kanonizálódó kérdéseit vizsgálja, és kritikai elemzés alá veti a jelenlegi normatív kereteket. Külön értékelendő a megmagyarázhatóság történeti áttekintése (9. fejezet), amely világossá teszi, hogy a „fekete doboz” problematika nem pusztán technikai, hanem mélyen episztemológiai kérdés.

Fogalmi és filozófiai észrevételek

A következőkben néhány olyan pontot emelek ki, amelyek nem a dolgozat értékét csökkentik, hanem inkább annak továbbgondolására ösztönöznek.

1. Az etikai fókusz kérdése

A cím – *Az MI etika új hullámai* – első olvasásra kifejezetten etikai fókuszt sugall. A dolgozat tényleges tartalma azonban jóval szélesebb: ontológiai, episztemológiai és technológiatörténeti elemzésekre épül. Úgy vélem, hogy az etikai kérdések a munka egészében inkább egy mélyebb filozófiai alapozásra épülnek, mintsem önállóan lebegnének.

Felmerül a kérdés, hogy az etikai problémák nem feltételezik-e a vizsgált technológia ontológiai és episztemológiai státuszának előzetes tisztázását. Másként fogalmazva: vajon nem célszerű-e az etikai reflexiót egy olyan filozófiai alapozás utánra helyezni, amely világosan tisztázza, mivel is van dolgunk? A dolgozat ezt a munkát részben el is végzi, de a cím hangsúlya ezt kevésbé tükrözi.

2. A dekompozíció episztemológiai státusza

A dekompozíció episztemológiai funkciója nem teljesen világos számomra. A munkamegosztás alapján bizonyára igaz, hogy sok ember alkotja meg a MI-t és ezek egyike sem látja át az egészet. De ez episztemológiai szempontból

másodlagos ahhoz a szerző által is megfogalmazott helyes megállapításhoz képest, hogy attól mert egy objektumot mi alkotunk, nincs közelebbi episztemikus hozzáférésünk hozzá, mint a világ bármely más objektumához. A PVC molekulát az ember hozta létre – talán nincs is más PVC molekula az univerzumban –, de ettől még a kémikus kutathatja, és tapasztalat útján ismeri meg a PVC tulajdonságait. Más szóval, az artefaktumokra vonatkozó tudásunk episztemológiai státusza semmiben sem különbözik a világ más elemeire vonatkozó tudásunk episztemológiai státuszától – függetlenül attól, hogy a szóban forgó artefaktumot egy vagy sok ember hozza létre.

3. Technológiába zártság

Az 5. fejezetben tárgyalt „technológiába zártság” fontos és aktuális probléma. Ugyanakkor az eredeti „vendor lock-in” fogalom szakpolitikai jelentése és a szerző által adott, szélesebb társadalomelméleti értelmezés között jelentős különbség van.

Míg az eredeti fogalom valóban egy káros, a szabad piaci döntéseket korlátozó jelenségre utal, addig a szerző által leírt jelenség nem káros, hanem ellenkezőleg: az emberiség történetét alapvetően átható, a történelmi léptékű progresszió irányát értelmező dominanciáról van szó. A szerző helyesen idézi be Marxot ezen a ponton. A társadalmi lét materiális alapjainak, a termelési viszonyoknak – benne a technológiai ismeretekkel, a „tekné”-vel – domináns szerepe van a társadalmi felépítménnyel szemben, vagyis a kultúrával, vallással, jogi vagy morális rendszerrel szemben. A szerző azt írja: „A kérdés tétje nyilvánvalóan óriási, hiszen amennyiben léteznek olyan technológiák, amelyek döntően befolyásolják a társadalmi folyamatokat, ugyanakkor már túljutottak a kontrollálhatóság pontján, akkor az emberiségnek egészen más stratégiákra van szüksége, mint olyan technológiák esetén, amelyeket még befolyásolni tud.” (74. old.) A technológiai változások és a társadalmi struktúrák viszonya nem feltétlenül írható le kontrollvesztésként. Mi kontrollálna mit? A termelési viszonyokban végbement változásokat kontrollálná a társadalmi felépítmény valamely eleme – a jog, ideológia, morál? Vagyis a társadalmi lét alapjaiban történő tektonikus mozgást kontrollálnia kellene valaminek, amit éppen ezek a tektonikus mozgások kényszerítenek ki? – ahol persze ez a „kikényszerítés” egyszerűen annak következménye, hogy a szóban forgó változás 8 milliárd lokálisan cselekvő ember hétköznapi létviszonyaiba épül be.

A fogalom pontosabb differenciálása még világosabbá tenné a fejezet állításait.

4. Fizikalizmus és funkcionalizmus

A szerző az elmefilozófiai fizikalizmust – hallgatólagosan, de szisztematikusan – a funkcionalizmussal azonosítja. Ez azért problematikus, mert a fizikalizmus csupán annyit mond, hogy az elme minden jelensége, minden sajátossága értelmezhető egy szigorúan fizikalista ontológiai kereten belül. Az emberi agy ugyanúgy elhelyezhető a fizikalista ontológiában, mint egy komputer; egy biológiai neurális háló éppen úgy fizikai, mint egy mesterséges, pl. elektronikai természetű vagy komputeren szimulált neurális háló. És nincs semmiféle ontológiai ugrás – hiszen a fizikalista ontológiai doktrína szerint nincs is honnan hova ugrani – amikor ezek a fizikai objektumok az elme valamely sajátosságának tekintett jelenséget mutatnak. Ezzel szemben a funkcionalizmus minden formájában végül is platonista. Ontológiájában olyan absztrakt/platóni entitások létezését tételezi, amelyek magukban foglalják a tudat esszenciáját. Ezek az absztrakt entitások valahol a platóni entitások szférájában ülnek, függetlenül azoktól a fizikai „realizációktól”, „implementációktól”, mint az agyak vagy komputerrek, vagy bármi – a „bármi” nem túlzás, ha pl. Putnam trivialitási argumentumára gondolunk, legalábbis az absztrakt/matematikai komputáció fizikai implementációjának vonatkozásában.

A 11. és 12. fejezetekben nyújtott elemzések szempontjából lényeges lehetne a fogalmi különbségek explicitebb követése.

Összegzés

Héder Mihály értekezése jelentős és érett munka. A szerző nem csupán áttekinti az MI etika jelenlegi vitáit, hanem önálló fogalmi keretet kínál azok értelmezésére. A konstruktóri szemszög bevezetése, az episztemorális döntések elemzése és a műszaki tudás filozófiai megalapozása egyaránt komoly hozzájárulások.

A kritikai megjegyzések a dolgozat ambíciójából és gondolati gazdagságából fakadnak. A munka nem lezárt rendszer, hanem élő filozófiai vállalkozás, amely vitára hív.

Összességében megállapítható, hogy a disszertáció a témában nemzetközi mércével mérve is releváns, koncepcionálisan koherens és eredeti hozzájárulás.

A disszertációt nyilvános vitára bocsátásra javaslom.

Budapest, 2026 február 22.

E. Szabó László