СТАНОВИЩЕ

за дисертационния труд на Лина Георгиева
„Процесуално-онтологичен подход към проблема за изкуствения интелект“
от доц. д-р Росен Люцканов, Институт по философия и социология (БАН)

Дисертационният труд на Лина Георгиева е с обем 176 страници, включващи увод, четири глави, заключение, приложение и библиография с цитирани източници, възлизаща на 160 заглавия на български и английски език. Обект на изследването е (установяването на) връзка между теорията за изкуствения интелект и (едно от направленията в) приложната онтология. Негова цел е да покаже, използвайки средствата на процесуалната философия, че „проблематиката на изкуствения интелект има своя потенциал и в полето на философията“ (с. 4). Тази формулировка създава впечатлението, че работата ще прилага подходи от теорията за изкуствения интелект в полето на философията, въпреки че изложените по-долу задачи на изследването (пак там) показват, че идеята е по-скоро обратната.

Първа глава („Исторически преглед на идеята за изкуствен интелект“) има характера на концептуално въведение в темата. Тя започва с уместната констатация, че съществуват различни типове интелигентност и че интелигентността е сложна способност, съставена от относително независими компоненти (с. 8). Раздел 1.1 разглежда различни митологични и литературни източници, разказващи за „същества, притежаващи, и даже надхвърлящи, възможностите на човешката мисъл“[1] (с. 10). Релевантността на всички представени примери изглежда съмнителна, доколкото голяма част от тях („чудните произведения“ на Хефест, „самоходните устройства“ на Архит и „пневматичните машини“ на Херон) говорят по-скоро за *автомати* – устройства, способни на самодвижение (с. 10-12), без претенция за пълнокръвно самостоятелно мислене. Същото може да се каже за „хуманоидните автоматизирани машини“ които представляват „имитация на човешки анатомични части“, каквито са били разработвани през Ренесанса (с.14). Прегледът всъщност показва, че идеята, че мисленето може да бъде пренесено в субстрат, различен от биологичния, е доста по-късна [повод да констатираме този факт ни дава и кратката бележка относно терминологията, която е типична за различните епохи – вж. с .18-19]. Дори примерът с „калкулаторите“ на Паскал и Лайбниц не опровергава това твърдение – ако погледнем по-внимателно приведения цитат (с. 15) ще видим, че според самия Лайбниц всъщност смята не калкулаторът, а човекът, който използва този уред. По-пряка връзка с темата имат „механистичната философия“ и „френският материализъм“, чийто най-ярък представител безспорно е Ламетри (с. 16). Напълно съм съгласен, че опитът да видим в машината човешки черти (каквато е интелигентността) е корелативен на опита да видим човека като машина. По-нататък, стигайки до романтизма, Мери Шели и нейният „Франкенщайн“, съмнявам се, че тук имаме основания да говорим за „завръщане към античните идеи“, въпреки подзаглавието „Новият Прометей“ (с. 17). В случая препратката към Прометей цели да насочи вниманието ни към бедите, които следват от опита на човека да се намеси в природния ред чрез технологиите. Концептуалните и стилистични влияния върху „Франкенщайн“ са добре проучени и е трудно да се твърди, че в значителен обем могат да бъдат съотнесени с античната епоха.

Раздел 1.2 е посветен на съвременните разбирания за изкуствен интелект, чиято основа е положена от теоретичен модел, станал известен по-късно като „машина на Тюринг“. В тази връзка е отбелязано, че „универсалната машина на Тюринг ... се различава от машината на Тюринг и двете понятия не трябва да бъдат бъркани“ (с. 21). Съгласен съм, че двете понятия следва да се различават, но универсалната машина на Тюринг е *вид* машина на Тюринг – такава, която може да симулира действието на *всяка* друга машина на Тюринг. По-нататък, правилно е отбелязан друг основен принос на Тюринг – в основата на предложения от него „тест“ (или „имитационна игра“) е положено разбирането, че интелигентността е аспект на

1 Запазил съм оригиналния правопис, свидетелстващ за обсесивно-компулсивно боравене с препинателните знаци, най-вече със запетайките.

поведението, а не проява на мистична нематериална същност, както твърди например картезианският дуализъм (с. 22-23). По-нататък Георгиева разглежда „възраженията“ срещу подхода на Тюринг, които той самият предвижда в оригиналната си статия от 1950 г. По повод на това се налага да отбележа още една неточност – първата теорема на Гьодел не твърди, че „всяка достатъчно силна математическа система ... съдържа твърдения, които не са доказуеми“ (с. 25). Това не е неочаквано, защото обикновено силно се надяваме нито едно неистинно твърдение да не бъде доказуемо в такава система. Проблемът е в това, че в нея могат да бъдат формулирани *истинни* твърдения, които не са доказуеми.

Раздел 1.3 представя исторически най-влиятелният опит за защита на ментализма – мисловният експеримент, предложен от Джон Сърл и станал известен под името „Китайската стая“. Тук също изложението съдържа моменти, които е добре да бъдат прецизирани: (1) за мен е неясно какво следва да означава „синтактичен смисъл“ (с. 30) (защото основната идея на Сърл е да различи смисъла на изказванията от синтактичните правила, по които боравим с тях), (2) какво е значението на твърдението, че „дигиталните компютри не могат да имат нищо повече от истинско разбиране“ (с. 31) (кой би могъл да има нещо повече от истинско разбиране и какво би било това „нещо повече“?), (3) защо за приведения по-долу цитат от Сърл се казва, че „отхвърля възможността за съществуването на силен изкуствен интелект“ (с. 31) при положение, че този цитат всъщност *представя* позицията на силния изкуствен интелект, (4) какво означава „тук откриваме разбиране не в системата, а в създаването на самото разбиране. Не е необходимо човекът в стаята да е този, който разбира, а да има разбиране като цяло“ (с. 33)? За мен е любопитно най-вече какво се има предвид с фразата „да има разбиране като цяло“. Така или иначе, напълно приемам обобщения извод от дискусията по темата – че интелигентността е продукт на една сложна система, която взаимодейства по сложен начин със сложната среда, в която е поместена, което извежда на преден план значението на влагането на изкуствения интелект в роботизирано тяло, което позволява такова сложно взаимодействие със среда, различна от виртуалната (с. 33).

Раздел 1.4 („Поглед в бъдещето“) представя накратко различни идеи за тенденциите в развитието на изкуствения интелект, а раздел 1.5 разглежда различни дефиниции на ключовото понятие за „интелигентност“. Тук също бих констатирал наличие на проблем, който не е анализиран в достатъчна степен. Георгиева заявява, че в своята работа ще се придържа към разбирането, че интелигентността се проявява в различни умения и способности, които не са особено тясно свързани една с друга, установяват се по различен начин при различни обстоятелства и се преценяват по различни критерии (с. 42). Това означава, че понятието за интелект е израз на ограничено приложима теоретична идеализация. Защо тогава в обсъждането на интелигентността в текста не е приложен по-нюансиран подход, който отчита всички тези основателни уточнения и се отказва да говори просто за (естествена или изкуствена) интелигентност? Според мен тук е пропусната прекрасна възможност темата да бъде развита в неочаквана и нетривиална посока. По-нататък, множественият характер на интелигентността може би оправдава говоренето за „идиоти-учени (саванти?)“, но не и за „протежета и други изключителни хора“ (с. 43) – думата „протеже“ обикновено подсказва всичко друго, но не и изключителност[2]. Главата завършва с констатацията, че „тестът на Тюринг е вече преминат“ (с. 47). Би било добре това твърдение да бъде подкрепено със съответна препратка, както и с уточнението, че са предложени различни тестове на Тюринг, на които са подлагани качествено различни системи – в крайна сметка „captcha“ също е тест на Тюринг, но от неговото преминаване (или не-преминаване) едва ли могат да се правят кой знае колко съществени изводи.

Втора глава („Приложна онтология и процесуален подход“) започва с разглеждане на отношението между приложно и теоретично знание и е призвана да защити твърдението, че „разграничението съвсем не е толкова лесно, ако не невъзможно“ (с. 52). Първоначално изложението има историко-философски характер и на моменти можем да съжаляваме, че

2 Мога само да предполагам, че тук се има предвид не „протеже“, а „prodigy“ - дарование.

дисертантката е пропуснала да разработи и аргументира част от приведените попътно тези: например, че при Сократ философията добива приложен характер (с. 53), или че трудовете на Аристотел са били „вдъхновение и основа за провеждането на експерименти“ и са „тласнали“ развитието на динамиката и механиката (с. 54). По-нататък, патосът на тази част от изложението остана неясен за мен, което ме кара да си поставя редица въпроси. Ако приемем заедно с Георгиева, че „всяко знание, в известен смисъл е приложна философия“ (с. 60), то какви дескриптивни функции би могъл да има етикетът „приложна философия“? По-нататък, ако прегърнем процесуалната философия (а с това приемем тезата за първичността на процесите) и отхвърлим метафизиката на субстанцията (която „схваща реалността като изпълнена със статични единици“), както ни се препоръчва в раздел 2.2 (с. 62), как това би ни помогнало при разработването на изкуствен интелект (още повече при положение, че това е теория, която изначално се интересува от изчислителни *процеси*, независимо от конкретния материален субстрат, в който те са били реализирани)? По-нататък, как „органичната форма на телеологично обяснение“, която има фундаментално значение за процесуалната философия (с. 66), се съотнася с онова, от което се интересува теорията на изкуствения интелект? Какво въобще тук подлежи на обяснение, позоваващо се на целеви причини? Накрая, дали обстоятелството, че дадена теория (подобно на квантовата физика) се интересува от „непрекъснато движещи се неща“ наистина е достатъчно да привидим в нея „наличието на процесуалния подход“ (с. 83)? Няма ли и други онтологични теории, които постулират фундаментално динамични първични същности и с какво точно ги превъзхожда теорията на Уайтхед? Накрая, как би изглеждало едно „емпирично потвърждение“, способно да „валидизира“ системата на Уайтхед (с. 88) и това ли е начинът за “валидизиране“ на една метафизична система? Не смятам за наложително на тези въпроси да се отговори веднага, но според мен обмислянето им е от първостепенно значение, ако дисертантката смята да продължи тази линия на изследване в бъдещата си работа.

Трета глава („Невронни мрежи, философия на съзнанието и процесуално-онтологичен подход към изкуствения интелект“) цели да отговори на три въпроса: (1) как се осъществява възприятието в една машина? (2) как тя би могла да се сдобие с емоционална интелигентност? (3) как се интегрират символните системи с външния свят? (с. 91). Тези въпроси безспорно са релевантни спрямо обсъждания проблем, тъй като, както беше показано в първа глава, интелигентността не е монолитна, а представлява съвкупност от различни способности за обработка на информация. В рамките на разглеждането на невронните мрежи са обсъдени два вида конекционизъм - „елиминационен“ и „имплементационен“ (с. 96), както и различията между системите от отворен и затворен тип (с. 97). След това изложението взима донякъде неочакван обрат – преминава към излагане на доктрината на панпсихизма. Както тя се разбира днес, тази доктрина се свежда до тезата, че „всеки материален обект има части, притежаващи ментални свойства“ (или, както е казано в текста, „съзнанието е фундаментално и вездесъщо“). Ако приемем, че това е така, то проблемът, с който се бори теорията на изкуствения интелект (как да създадем материално устройство, което притежава ментални свойства, каквато е интелигентността), просто *изчезва*. Няма проблем да създадем такива материални устройства, защото цялата материя по начало притежава ментални свойства. Искрено се съмнявам, че това осигурява полезни и практически приложими препоръки на онези учени, които в момента разработват подобни устройства. Съответно, разглеждан от такава гледна точка, „панпротопсихизмът“ на Уайтхед, който ни препоръчва разбирането, че „елементарните форми на материя, могат да бъдат свързани с елементарни форми на преживяване“ (с. 116) трудно може да бъде признат за „сериозно занимание“, както Георгиева сама отбелязва (с. 117). Доколко съотнасянето му с хипотетичната „обективна редукция“, за която говори Пенроуз, може да оправдае по-сериозното отнасяне към тази теория (с. 118) е въпрос, на който не мога да отговоря. По-нататък в същия раздел са обсъдени пределно накратко въпросите за свободната воля и личностната идентичност, които са свързани с темата за съзнанието. В тази връзка е направено интересното наблюдение, че в следствие от пребиваването ни във виртуална

реалност личностната ни идентичност се размива извън границите на нашето тяло – ние сме своите „аватари“ в същата степен, в която сме собственото си тяло (с. 107). Накрая, има един момент в аргументацията, на който бих искал да обърна специално внимание. Георгиева представя тезата на Франц Риферт (мотивирана чрез „процесуално-онтологична методология“), според която „човешкото мислене, е изключително зависимо от контекста“ (с. 119). Тази теза е „потвърдена“ чрез следния „емпиричен експеримент“: събрани са експериментална и контролна група, при чието съставяне са използвани „онтологичните схващания на Уайтхед за актуално биващи“, т.е. „те не са нещо статично, а силно динамични“. След това контролната група получава „писмени *о*казания“, а експерименталната – същите инструкции плюс „неволно подсказване“. Оказва се, че делът на верните отговори е по-висок при експерименталната група. В тази връзка имам няколко въпроса: (1) В какъв смисъл участниците в групата са „силно динамични“? (2) Какво означава един експеримент да бъде проведен съгласно „процесуално-онтологична методология“? (3) Нужно ли е да се позоваваме на Уайтхед, за да стигнем до извода, че човешкото знание е „контекстно зависимо“, или има и други пътища, по които можем да установим този отдавна известен факт? (4) Следва ли да сме изненадани, че онези, които получават подсказване (макар и „неволно“) дават повече верни отговори? (5) В светлината на казаното, можем ли да разглеждаме този експеримент като потвърждение за „базовото философско предполагане на Уайтхед“ (пак там)?

Четвърта глава („Изкуственият интелект в сферата на образованието като процесуално-онтологичен подход“) обсъжда настоящото състояние и перспективите пред използването на изкуствен интелект в образователната сфера. Ясно са показани предимствата на изкуствения интелект в тази сфера, които обуславят широкото му навлизане посредством различни подходи (с. 122ff). От друга страна са посочени и недостатъци, които обуславят приемането на тезата, че „ние се учим най-добре от практикуването с други човешки същества“ (с. 127). По-нататък, показано е, че процесуално-онтологическият подход обуславя приложения с по-скоро трансдисциплинарен (снемащ разликите между различните дисциплини), отколкото мултидисциплинарен (обединяващ механично различни дисциплини) характер (с. 128). Тук срещаме и едно интересно наблюдение: понякога проявяваме повече търпение към машината (когато я подлагаме на машинно обучение), отколкото към обучаващите се деца (с. 131). Това наистина е парадоксално – причината вероятно е в това, че поначало очакванията ни към машините са силно занижени (когато станем по-взискателни към тях в резултат от подобряването на общото качество на представянето им вероятно това вече няма да бъде така). Съвсем основателно е посочено също, че обучението е двупосочен процес, при който учениците се учат от учителя, но и учителят (независимо дали е човек или не) се учи от учениците (с. 134, 140). Заключителната част, която изтъква ползите от видео-игрите в процеса на обучение, според мен също повдига редица въпроси. Първо, очевидно е, че днешните младежи на възраст между 8 и 18 години прекарват все повече в игра на електронни устройства (с. 143). Можем да допуснем също, че момичетата, която играят видео-игри, е далеч по-вероятно да се ориентират към обучение в STEM-специалности (с. 143-144). Изглежда съмнително обаче, че играното на игри само по себе си е причина за това и дори е в състояние да *създава* нужната мотивация (с. 144) – в случая отношението между причина и следствие може да е точно обратното: предшестващият интерес към наука и технологии може да е причина за повишена склонност да прекарваме повече време пред компютъра (в игра или учене). Допълнително изясняване изисква също идеята за разработване на специална видео-игра, в която да бъдат заложени идеите и достиженията на процесуално-онтологичния подход (с. 147). Според Георгиева, тази игра ще ни позволява да оценим времето, което учениците отделят на различни аспекти на образователното съдържание. Очевидно обаче, причините за това може да са различни – повече време отделяме или на това, което ни е по-интересно, или на онова, което разбираме по-трудно. За да се отстрани тази неяснота се предвижда използване на въпросник (с. 147-148). Тогава обаче остава неясно какво точно осигурява играта и не можем ли да се задоволим с

въпросния въпросник (една далеч по-скучна, но и далеч по-добре проучена възможност). Всъщност, не можах да си изясня какво точно се предвижда да бъде отношението между играта и въпросника – в тази връзка е посочено само, че „самата игра ще се явява под формата на въпросника“ (пак там). Струва ми се, че една игра, която се явява под формата на въпросник може със същия успех да се определи просто като ... въпросник.

В заключение мога да отбележа, че дисертацията засяга изключително широк кръг от въпроси. Това само по себе си не винаги позволява те да бъдат обсъдени достатъчно подробно и задълбочено. Съответно, на места в предложенията за прилагане на процесуално-онтологичен подход в сферата на изкуствения интелект и образованието липсва конкретност, дискусията като цяло остава на абстрактно-методологично ниво, което следва да се има предвид при оценяването на заявените в авторефератa приноси. Въпреки това, имайки предвид безспорната значимост на темата и евристичността на използвания подход, ще гласувам за това на Лина Георгиева да бъде присъдена образователната и научна степен „доктор“ по направление 2.3 (философия).

08.03.2020,

София /Р. Люцканов/