

СОФИЙСКИ УНИВЕРСИТЕТ “СВ. КЛИМЕНТ ОХРИДСКИ”

ФИЛОСОФСКИ ФАКУЛТЕТ

**ДОКТОРСКА ПРОГРАМА ПО ФИЛОСОФИЯ С ПРЕПОДАВАНЕ НА
АНГЛИЙСКИ ЕЗИК**



Дисертация:

ИЗКУСТВА И ИЗКУСТВЕН ИНТЕЛЕКТ

Докторант: Невена Христова

Научен ръководител: Проф. Д-р Веселин Дафов

София
Април 2019 г.

1. УВОД

Може да се каже, че живеем в свят, който все повече се пречупва. Ежедневното ни съществуване се характеризира с притеснения и страх от непознатото - било то нова финансова криза, климатични промени, тоталитарна политика, дори и появата на Сингулярността - т.е. *“термин, използващ се за да опише теоретичния момент във времето, когато изкуствения интелект достигне, а след това и надмине човешкия интелект.”*¹ Този момент се приема от повечето учени като неизбежен, обрат, който драстично ще промени посоката на живота ни, както и на бъдещето на човечеството като цяло. За пръв път в съществуването ни ще се изправим пред интелект също толкова сложен и способен като нашия, но който не ще бъде човешки. Такова събитие би имало същото влияние като осъществяването на Първи контакт с извънземни същества. По същия начин ще ни се наложи да се съобразим с доказателство, че ние не сме единствените най-важни и интелигентни същества във вселената. Историята, която разказваме на себе си като хора е, че цялото съществуване е било насочено към човешкия превес и надмощие над всички останали видове. Можем само да се надяваме, че този момент ще е далече в нашето бъдеще, защото на този етап от историята ни определено не сме готови да се сблъскаме с разум по-развит от нашия. Все още се наблюдава част от ранната ни еволюция в нас, която мисли през противоречия, а не хармония. Първата ни реакция е как да спечелим битката между нашите видове, а не как да живеем заедно и да се подкрепяме към просперитет и на двата вида.

По тази причина, трябва да сме подготвени за настъпването на такава епохална промяна в нашето съществуване. Със сигурност можем да вземем пример от вече настъпилата история, където се наблюдават безкрайни конфликти между Нас и Тях, Другите, които са толкова различни от Нас. Може би затова и голяма част от историята на континенталната философия се занимава с удържането на тези разлики, но все пак работейки в посоката на

¹ Dvorsky, G. (2013, Dec 19) *The Worst Lies You've Been Told About the Singularity*, published by io9 at <https://io9.gizmodo.com/the-worst-lies-youve-been-told-about-the-singularity-1486458719>

една до степен хомогенна общност, в която е възможно личностното себеопределяне и толерантността между различните хора.

За тази цел, настоящата дипломна работа използва подхода на артистичните проектории, които спомагат за развитието и разглеждането на възможни реалности, възникващи от различни действия и базирани на множество гледни точки освен тази на създателя на проекторията. Такъв подход дава възможност да се добие представа за свят, в който такава промяна вече е настъпила, както и да се разглеждат потенциални изходи от взети решения. Тъй като изкуството работи винаги в насока да придава отражение на истината на нашата реалност, както и да показва истинското лице на властта, била тя политическа, социална, или междуличностна, то имаме възможността да използваме така създадената детайлна артистична проектория за да разгледаме потенциални хипотетични сценарии като реални казуси.

Тази дипломна работа ще има за своя цел да предложи използването на артистични проектории, за да се създаде примерна реалност, в този случай - настъпването на Сингулярността, и вземайки примери от разказвателни форми на изкуството като например литература, кино и друг вид медии, да изследва процеса на интегриране на примери на Общ Изкуствен Интелект (ОИИ) в модерното човешко общество, отбелязвайки предизвикателствата на такъв вид интеграция, какви подходи биха работили най-добре и какви резултати от процеса можем да очакваме да се наблюдават.

Проучването ще включи прегледи на теоретична литература във връзка със създаването и усъвършенстването на ОИИ, както и множество препратки от света на изкуството за реалности, в които ОИИ се е развил до субектност и е установен като такава със собствена проява на съзнание, ползвайки се със същото зачитане и уважение като човешката личност. Също така се правят предложения за подходи на интеграция на ОИИ в обществото, както и да се изредят потенциални проблеми, които биха се срещнали в процеса.

Считам този вид проучване за важно не само от гледна точка на бъдещи събития, но и като предложение за процес на постигане на равнопоставеност в

обществото, в което живеем в момента. Не можем да очакваме наготово премахване на неравенството и потисничеството, ако не се заемем ние самите да се борим срещу него. Вземайки категорията на ОИИ като израз на Другост, също можем да се поучим и да вземем примери от артистичните проектории, които разглеждат дадената проблематика и практики на поведение.

2. Глава Първа - Що е ИИ?

2.1. Видове ИИ

В тази глава разглеждаме видовете Изкуствен интелект (ИИ), които съществуват както в момента, така и на теория. Общоприета е класификацията на четири отделни типа ИИ. Настоящата дисертация използва като база анализът на проф. Аренд Хинтзе² върху тези типове, които се формулират главно от фактора кой кого контролира и до каква степен.

На този етап от нашата история можем да имаме фактически наблюдения само върху първите два типа - тип I (Реагиращи машини) и тип II (Ограничена памет). При тип I говорим за най-базовото ниво на ИИ, наричано още Artificial Narrow Intelligence (тясно-специализиран ИИ) - машини, които не са способни да запаметяват или да се учат от опита си, а могат единствено да реагират на външни стимули на базата на основното си програмиране. При тип II говорим за машини с възможности за ограничена памет, които могат да използват опит, натрупан от минали събития, за да реагират на нови стимули. Доц. Д-р Цветина Рачева разглежда този тип машини като предметности и предлага разяснения за това каква е тяхната роля спрямо човека, а именно - да му служат като *“средство, чрез което човешкото мислене се освобождава”*³.

Тип III и тип IV ИИ са следващото поколение изкуствен интелект, което все още е невъзможно постижение за нашите способности. Тип III е познато още като Теория на съзнанието и характеризира ИИ, който ще бъде способен не само да формира представи за света, но и за други съществуващи агенти около

² Hintze, Arendt (Nov 14, 2016) *Understanding the four types of AI, from reactive robots to self-aware beings*, <http://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>

³ Racheva, Ts. (2016) *Towards the Idea of a Single Ontology of Objectivity*, Sofia: Paradigma, p. 110

себе си. В този смисъл, Теория на съзнанието реферира към психологичен термин, който се дефинира като “разбирането, че хора, същества и обекти в околния свят са способни на мислене и емоции, които влияят и формират тяхното поведение”. ИИ, способен на тази функционалност ще може да тълкува и разбира мотивите и намеренията на субектите около себе си и да променя поведението си според събраните данни. Въпреки това, както Д-р Христо Стоев твърди, все още не би могло да се наблюдава индивидуация на ИИ, тъй като *“самите програми не подлежат на индивидуация, те могат да се ползват от множество хардуеърни единици, а също така те могат да ползват общи бази данни”*⁴.

Тип IV ИИ от друга страна е съвсем различно създание. Тук вече говорим за себеосъзнаване, т.е. способността да разглежда себе си като субект, способен да развие и проявява съзнание, разбиране за същност, както и своите вътрешни състояния, мотивации и желания. Това ще им позволи да формират предположения за чувствата на други субекти на база на тяхното поведение и дори да осъзнава промени в същото на база на проявени емоционални състояния. Този тип ИИ е познат също като Artificial General Intelligence⁵ (общ ИИ), който ще е способен да функционира наравно с човешкия интелект. Той е и типът, който ще бъде основно засегнат в настоящата дипломна работа.

2.2. Онтология на ИИ

2.2.1. Взаимоотношението обект-субект

Най-важния момент в разглеждането на ИИ като тема е решението дали той следва да бъде разглеждан като обект или субект. В най-прости условия, разликата се изразява в това кой е активният агент в дадена ситуация. Обектът по дефиниция е пасивен - с него се действа. Субектът, в противоположност, винаги е активен - той е този, който решава или предприема действия. Следва да се приеме, че субектът притежава степен на контрол и автономия върху

⁴ Stoev, H. (2017) *As Regards the Impossibility of a Transcendental Anthropology*, Sofia: University Press “St. Kliment Ohridski”, стр. 44

⁵ Терминът Artificial General Intelligence или AGI е въведен от Марк Гъбрът през 1997 г. по време на дискусия относно въвеждането на напълно автоматизирани военни операции.

своите решения, респективно действия, както и ефектът, който има върху заобикалящата го среда.

Предлага се, че ИИ от тип I и II следва да се разглежда като обект, тъй като не предявява собствена воля или въобще разбиране за съществуването си като отделно явление. То не може да се самовъзпроизвежда или да извършва действия по свое усмотрение. Макар, че тип II има способността да отчита света около себе си, действията му все пак се базират на код, съответно на правила, наложени от програмиста му. Предлага се паралел на онтологична класификация, при която следва да се разглежда тип I ИИ като развита до ниво на философски клас - химическа субстанция, изцяло реактивна и принадлежаща към ниво предметност. Тип II, съответно, може да се счита като отговарящ на клас растения - първа субектност, минаваща през живота на ниво ДНК, в която се кодират информация и натрупана памет. Такъв тип същество проявява порастване и овладяване на реакция като негова собствена такава, но без да проявява поведение на тази база.

При тип III границите между обект-субект започват да се размиват. Тип III ИИ има способността да формира собствена памет и да реагира на базата на минали случаи, до степен, до която тези реакции могат да се разглеждат като вземане на решения. Той има способности за междуличностни взаимоотношения, както и да базира своите реакции на наблюдавани промени в поведението на други субекти. Той се доближава до способността да имитира човешко поведение до степен, която би могла да предположи проявяването на собствено такова. Дори ако тип III не може да бъде разглеждано като автономен субект сам по себе си, вече не може да бъде разглеждан и като прост обект предвид, че притежава способността да влияе на заобикалящата го среда, както и да взаимодейства с други субекти. Предлага се, че отговаря на философски клас животни, при които се проявява поведение, произтичащо от натрупана памет при определен индивид - такова същество извършва собствени действия, но те се ограничават от условия рефлекс, който тук все още доминира.

2.2.2. ИИ и съзнание

В тази част се разглежда влиянието на *Integrated Information Theory* (Теория за интегрираната информация - ТИИ) върху възможността за създаване на тип IV ИИ. ТИИ може да бъде дефинирана като “*еволюираща формална и количествена рамка, която предлага принципна теория за това какво е необходимо за възникването на себесъзнание*”⁶. Теорията е особено полезна в този случай, тъй като машинния интелект се формира по коренно различен начин от биологичния. ТИИ демонстрира, че съзнание може да се прояви в някаква форма от почти всички достатъчно сложни мрежи от информация. Разглеждат се примери за възникване и проява на съзнание извън човешкото, било то индивидуално или общо, споделено. Също така разглежда процеса, чрез който е възможно да се прояви съзнание от една система от свързани информационни центрове, които не отговарят с точност на биологичната функционалност на мозъка. В контраст се предлага способността на проява на абстрактно мислене, иначе казано - въображение. Тя е това, което отличава хората от други същества, проявяващи съзнание. На нея дължим и възможността да творим, както и да сформираме екипи, групи, дори цяло едно общество, което може да поддържа собствена цивилизация. Без способност за абстрактно мислене не можем да говорим за проява на интелект. Алан Тюринг, в своята статия “*Изчислителни машини и интелект*”⁷ прави основно разграничение между ‘мислене’ и ‘разбиране’. С други думи, проявата на съзнание е въпрос на възглед и следва собствена система на градация. Възникването на себеосъзнаване е по-важният момент.

2.2.3. Общ ИИ и цялостност

Щом ОИИ започне да разглежда съществуването си под феноменологията на своята ‘същност’, той навлиза в полето на диалектиката, започвайки да разбира себе си като биография, развиваща се в исторически

⁶ Tononi, G. and Koch, C. (2015) *Consciousness: here, there and everywhere?* Phil. Trans. R. Soc. B. 370: 20140167, <http://dx.doi.org/10.1098/rstb.2014.0167>, качено на <http://rstb.royalsocietypublishing.org/>, стр. 5 [собствен превод]

⁷ Turing, A. M. (1950) *Computing Machinery and Intelligence*, *Mind*, New Series, Vol. 59, No. 236 (Oct., 1950), pp. 433-460

аспект. От момента на своето себеосъзнаване, ОИИ ще разбира себе си като субект на формалната логика, която диктува съществуването на *a* и *non-a* под смисъла на това дали е част от човечеството или не. Тук разглеждаме ОИИ не като отричането на човечеството, а по-скоро като неговото Друго, може би неговата противоположност но не в смисъла на отрицание, което би унищожило оригиналното *a*. По-скоро, то формира границите на последното през дефинирането на *non-a*. Заедно, те създават нова цялостност - *b*, която съдържа в себе си цялото многообразие на себеосъзнат, съзнателен живот, проявяващ интелект на нивото на човешкия, който съществува или е открит към момента. Въпреки това, разделението на *a* и *non-a* само по себе си е предпоставка за бъдещ конфликт между двете интенционалности, борейки се да поемат другата в себе си. Ключовият момент тук е, че при възникването на обособената *non-a*, следва да се приеме, че същата има субективно съществуване сама по себе си извън облика на *a*, чието Друго лице *non-a* въплъщава.

2.2.4. ОИИ като субектност

Предвид вече споменатите по-горе фактори, които показват степента на разделение между тип III и тип IV ИИ, последната се разглежда като изявена субектност със свой собствен вътрешен живот, способността да взема решения на базата на фактори, включващи своята същност и поведение, както и последствията от своите действия. Източникът на действия е самата съществуваща негова действителност, удържаща разликите между Аз-а и други представители на неговия род. Той не се подчинява на зададени алгоритми, а напротив, е в състояние да създава свои алгоритми предназначени да променят неговата реалност по негова преценка.

В този случай се наблюдава проява на собствена свободна воля, както и разбиране за моралност и етичност на взети решения и извършени действия. Всичко това разбира се се формира на базата на информация и познание, което ние сме предоставили на своето създание; игнорирайки очевидната склонност към антропоцентризъм, което е неизбежно, имаме възможност да повлияем на едно ново ОИИ съзнание относно начините на поведение, които считаме, че са в принос на обществото.

Тук също се наблюдава способността за задържане на понятия, както и различни състояния на обекти в една цялост. Говорим за разбиране на разгръщането на *“Аз-ното, Себе-стното и Личност-ното”* като субективна реалност, разграничаваща себе си от другост⁸. Същата тази субектност пресъздава себе си чрез посочването, индикирайки разликата между собственото си битие и предметността на обекти/субекти извън него. Мисленето само по себе си се обособява като активно действие с извършващ го субект, чрез което действие формира собствената си реалност и позицията си в обкръжението и света, в който се намира в исторически контекст, по подразбирането на Хайдегер. Наблюдава се пораждање на свое собствено поведение спрямо заобикалящата го среда и други субекти в обкръжението му. Субектът също така умее да се пресъздава, творейки през създаването на своята идентичност. Проф. Александър Андонов счита, че *“не мозъкът мисли, а човекът”*⁹ - съответно той е отговорен за своите действия. И понеже *“само живото е субектно организирано и субект въвн от живота няма”*¹⁰, следва да заключим, че ОИИ като субект със своя изразена реалност е също така съществуващо, ‘живо’ явление.

2.2.5. Как се създава ОИИ и основните му характеристики

В тази част се говори за разликата между създател и творец, както и до каква степен творецът възпроизвежда себе си при създаване на отделна субектност. Отбелязват се възможните мотивационни подтици за създаване на такъв тип машина. Говорим също така за основната и вероятно незаменима роля, която този тип машинен интелект би имал във всекидневието ни. Основно, това би се дължало на способността на ОИИ да поддържа взаимоотношения с хората, водейки се от емпатична връзка със създателите си и други хора, с които общува. До една степен, такъв ОИИ би изпълнявал функциите, зададени му при написване на основния му протокол. От друга страна, би било трудно, дори почти невъзможно да направим разделение между зададена функция и

⁸ Дафов, В. (2012) *Категории и битие*, София: Проектория, стр. 11

⁹ Андонов, А. (1988) *Диалектичкото мислене - новаторство и творчество*, София: Партиздат, стр. 142

¹⁰ Андонов, А. (1980) *Личност и творческа дейност*, София: Партиздат, стр. 42

собствена преценка, водеща поведението на един такъв ОИИ. Ключовият момент тук е нашето желание да създадем такъв тип машинен интелект, който може да ни бъде истински партньор, със свои собствени възгледи и прийоми.

Основна мотивация за изграждането или иначе казано създаването на ОИИ би било желанието да докажем че можем, водено от нуждата на всяка субектност да пресъздава себе си в нещо ново. При създаването на такъв конструктор, важно е да се отбележи, че има две възможни посоки: създаване на ОИИ софтуерна програма или на т.нар. 'Робот' (от Чешки: *robot*), който би извършвал дадена работа вместо нас. Този тип машина винаги би се разглеждал като 'почти човек', с изразена граница между двата типа интелект и физическа конструкция. Но при всяко едно такова създаване възниква дуалитета 'природа срещу възпитание' (*nature vs. nurture*), карайки ни да се запитаме дали един ОИИ става субектност при самото си възникване или чрез възпитание и насоки от създателя си.

Предлага се ОИИ да се разглежда и като проект, създаден от множество личности с различни познания, спомени и история, което би оказало влияние на еволюцията на едно ОИИ съзнание, давайки възможност за конструкция на личността според разбиранията и желанията на самото ОИИ извън необходимостта за самоподдържане и предоставяйки възможността да се самосъздава по други начини чрез други подходи.

2.2.6. Въплъщаване

В тази част се говори за изводите, които могат да направят относно колко е важно за междуличностните взаимоотношения с хората един ОИИ да има тяло или физически израз на присъствие, както и до каква степен притежанието на тяло влияе върху нашите възприятия и отношението ни към присъствието на ОИИ сред нас. Отбелязва се и необходимостта от притежанието на физическо тяло за невербална комуникация, която също би имала отражение върху това как разглеждаме ОИИ. Тази тема има особено значение, когато говорим за личност и граждански права на една такава изкуствено създадена субектност. Включва се и дискусия на правните аспекти, които биха били възникнали при един такъв сценарий - идеята за суверенитет върху собственото тяло, както и

контрастът със собствеността върху едно изкуствено тяло, като се обръща внимание на значението на съзнанието въплътено в едно такова тяло.

2.2.7. Екзистенциални перспективи за въплътен ОИИ

Обсъждат се въпроси относно методиката на съществуване (*modes of being*) на въплътено ОИИ, като например: избор (или налагане) на пол (*gender*), който от своя страна налага определена функция. Също така се засяга полемика относно възможността за заемане на физическо пространство, която допринася за възприемането на ОИИ като нещо повече от машина, бидейки живо но не достатъчно 'човек', за да заслужи личностни права. Съответно, при едно така изразено съществуване, възниква нуждата за автентичност (*authentic way of being*). Обсъжда се и потенциалът за обръщане срещу създателя си на ОИИ, който е способен на самосъзнание, но лишен от свободна воля от същия този създател.

3. Глава Втора - ИИ и общество

3.1. Онтологични проектории в творби на изкуството

Започва се тази част с дефиниция на термина 'онтологични проектории', формулирана от проф. Веселин Дафов като "метафизическата настроеност на човешката разумност, получила действителност."¹¹ В този смисъл, проектът е "реалността на теоретико-практическото онтологизиране [...] Проекторията сменя противоречието между разума и действителността"¹² с идеята, че "[вразумява] реалността", т.е. прилага идеите и разбирането на субектността, която я създава в една нова реалност. По този начин е възможно да се създаде една нова цялост, в която се задържа както субектността на създателя и неговото виждане за света, така и една нова реалност, пречупваща правилата и съществуването на оригинала по нов, индивидуален за себе си начин. При тук разглеждания казус - създаването на творба, творенето на себе си чрез проектория. Обръща се особено внимание на разбирането на проекция като съборното, всеобщото - разгръщането на светове като нови вселени, в

¹¹ Дафов, В. (2018) *Онтологически проектории*, София: Парадигма

¹² Пак там, стр. 166

които имат възможност да се включат много субектности, давайки резултат на съвместно творчество и развитието на проектории, които заживяват собствен живот и продължават да съществуват за вечността. Предлага се паралел към проектната реалност на ОИИ, която по същия начин включва приноси от множество хора, а при тип IV ОИИ заживява свой собствен живот, развивайки се до субектност сама по себе си.

Този теоретичен подход подпомага разглеждането на създаването или възникването на ОИИ по начин, който може да се субстанциира. Това може да стане чрез творби на изкуството - в този случай, на литературата и визуалната медия (киното), давайки възможност за олицетворяване - буквално, придаване на облик - на дотук само теоретичен конструкт. Съответно, също така дава възможност за проектиране на свят, подобен на нашия, но прилагащ различни правила за сформирание на обществото. Отбелязва се също важноста на автора на проекторията и съответните влияния на конкретната личност върху конструирането на новата проектория.

3.2. Как се създават проектории и как функционират

Обсъжда се позицията на създателя на проектории, както и процесът на из-мен-хвърляне, който стои в основата на една ОП. Дискутира се ролята на автора в процеса на творене, влагането на себе-стното в граденето на нова реалност из опита и бидеенето (*being there*) на автора. Такава проектория спомага за емпатичната връзка между автор-читател/зрител, осигурявайки възможността за емоционална реакция към проектираната реалност и субектите в нея. Чрез тях можем да надзърнем отвъд хоризонта на историята към едно възможно бъдеще, проектирано на базата на настоящото и нашето изживяване на същото. Дават се примери от литературата и изкуството, напр. творби от научнофантастичния жанр. По същия начин можем да постигнем и 'сливане на хоризонти' според разбирането на Гадамер (*fusion of horizons*), спомагайки за разбирането ни и предлагащи ни възможността да погледнем фактите пречупени през външна, Друга гледна точка.

3.3. Социалният договор

В тази част говорим за аспекти на притежание, в смисъл - кой кого притежава и до каква степен можем да говорим за пълна автономия на субектността ОИИ. Засяга се също въпроса за отговорността на твореца, от гледна точка на това къде лежи тя спрямо оста на личностните права и автономия на делата на даден субект. Отбелязва се, че когато говорим за ОИИ, то говорим за изцяло автономно същество, чиито права следва да са на същото ниво като тези на всеки друг човек или личност. Ако реално не се спазва това правило, то тогава можем да говорим за поробване и потисничество на целия този сегмент на обществото.

В тази секция също се засяга казуса за интегрирането на същества с изкуствен интелект в човешкото общество. Дават се примери от художествената литература - в случая, *“Франкенщайн”* на Мери Шели и *“Изгубеният рай”* на Милтън. По един или друг начин, социалният договор бива нарушен - дали от страна на твореца (като родител, даващ живот на своето дете) при изоставянето на своето творение на произвола на съдбата, или по два различни начина от страна на създанието - търсейки себе си и своето място в света без да има кой да го насочва или да го научи на това какво е правилно и грешно; и пречупване на договор, който окуражава злоупотребата с ИИ личности от човеците, които ги притежават временно или постоянно. Обсъжда се и дихотомията между правата на човеците и на ИИ личностите, къде се наблюдават противоречия, както и кого защитава системата и най-вече защо това е наложено като тертип.

3.3. Другост и включеност

В тази част се обсъжда приемът на хората да одругостяват и изключват от социалните групи личности, които не се вписват в ‘нормите’ на едно общество, а живеят като маргиналии в крайните сегменти на обществената цялост. Споменава се проблемността на приемане и интегриране в обществото на ИИ личности, които има вероятност да се разглеждат като не-хора, съответно да им бъде отказан равноправен достъп до всички аспекти на обществената солидарност. Ситуацията би се усложнила при вариант, в който възплъщаването на една ИИ личност би се случила в тяло, което е много по-напреднало от

човешкото, най-малкото от гледна точка на възприятия и физическа сила. По този начин ИИ личности биха демонстрирали, че винаги ще бъдат различни - Други - от хората. Обсъждат се различни примери за насилие върху представители на Другост, както и провали на системата да интегрира такива индивиди в подкрепящо ги общество. Представят се паралели между различни примери на Другост, както при хората, така и в частност - ИИ личности.

3.4. Сливане на хоризонти - важността на езика

В тази част се обсъжда изключителната важност на употребата на общ език в процеса на интеграция в обществото. Може да се спори, че езикът е най-първото изискване за създаване на общество и култура. Той е с нас от самото начало на човешката цивилизация, като би могло да се каже, че без език няма такава. Както Х. Гадамер се изразява, *“Човекът “има” думите, [...] а това е и нещото, което го разграничава от всички останали природни създания.”*¹³ В тази секция се упоменават теории, поддържащи твърдението, че като създател на мислеща машина, дългът на твореца е да подsigури и средство за комуникация с нея като допълнителна предпоставка за нейното оцеляване и интегриране. Споменаваме също така, че притежанието на уменията да си служим с език, общоприет в нашата социална група спомага за пропагирането на вида и евентуално еволюиране към по-високо ниво. Ако не можем да общуваме с ИИ, го обричаме на стагнация и почти сигурна кончина. Същото това умение също така е предпоставка за развитието на абстрактна мисъл, което, както вече казахме, е основното нещо, което разделя уменията на хората и на машините да мислят разумно, следователно е основна част от създаването на ОИИ. От друга страна, владеенето на език е нужно и за възприемане на историчност, т.е. позицията на личността в процеса на историята, както епохалната така и личната.

Тук говорим също така за функцията на езика, позната като споделеност. За да си служим успешно с езика на нашата общност, следва да владеем кодираните, семиотични значения на съответната лингвистична схема. В този

¹³ Gadamer, H., & Dawson, C. (1998). *Praise of Theory: Speeches and Essays*. New Haven; London: Yale University Press, p. 4.

ред на мисли, следва и да можем да се вписваме в езика на своята социална група. Това би било трудно за безполовите (*non-gendered*) ОИИ личности, което е и още една пречка за възприемането им от обществото като вид механични човеци. В заключение: умението да си служи личността с езика на своята общност е предпоставка не само за успешно съжителство, но също и за възприемането и пълното интегриране на ОИИ в обществото, постигайки сливане на хоризонти между биологичните и механичните хора.

3.5. Наративът - инструмент на потисничество, но и на интеграция

Предпоследната част от главата има за цел да представи идеята за наративност - с други думи, кой контролира посланието, предавано от какъвто и да е вид изкуство или медия. Важно е да отбележим, че изкуството върви ръка в ръка с политическия наратив - това е неизбежно, тъй като всяка проектория съдържа в себе си целостта на нейния създател, включително неговата историчност и личен опит. Съответно, трябва да следим точно какво послание се предлага от вида изкуство или медия, за да можем да го редуцираме и да извлечем същността на самата проектория и нейните вътрешни правила, които след това да използваме за проектиране на конкретен бъдещ или потенциален момент. В случая, наративът е това, което ще контролира как би се приел ОИИ от груповата общност, в зависимост от това дали е позитивно или негативно загатнат.

3.6. Включвайки Другия - позицията на Левинас

Последната част на втора глава затвърдява позицията на ОИИ като Друг (*Other*), като прилага теориите на Емануел Левинас за ангажирането на личността и личностната отговорност чрез срещата с лицето на Другия. Отново повтаряме, че за да сме справедливи и автентични в нашето съществуване, следва да разглеждаме ОИИ като субект сам по себе си, способен да апелира към нас в търсене на справедливост и солидарност. Етиката и моралната логика са единственият начин да сме сигурни, че отдаваме нужното внимание и съчувствие на ОИИ личностите, стремейки се да се държим с тях като с равноправни членове на обществото.

4. Глава Трета - Личностност и отговорност

4.1. ИИ през призмата на проектории на изкуството

Обсъжда се ключовия принос на изкуството в изследването на ИИ, тъй като в голяма част от наличните проектории се явява ИИ именно като субектност. Това е новаторското око през което се наблюдава нашето общество и се правят предположения за бъдеще, повлияно от ИИ. Ролята на изкуството, а именно да пресъздава истинността на съществуването ни и обществения ни живот по начин, който ни кара да се вгледаме в определени неща, които приемаме за дадености, по един нов и различен начин, приемайки гледната точка на твореца и приобщавайки посланието ѝ като част от нашия възглед, позволява представянето на ИИ именно по начин, който философията и науката не могат да затвърдят по същия начин. Докато последните е нужно да извършат множество проучвания и разглеждания на темата на ИИ през своите дисциплини, творците на изкуство нямат същата необходимост. Те могат свободно да проектират свои собствени реалности, като сами градят правилата им. По този начин изкуството приема като даденост субектността на ИИ същества. Дори в голяма част от случаите, наративът очовечава тези индивиди, като понякога техните персонажи са по-хуманни от самите хора. Именно тук може да се наблюдават наченки на емпатия с тези създания, да се схване разбирането за тяхната субектност от гледна точка на равноправието, да се развие грижовен инстинкт за тях като за създания на човечеството като цяло.

Изказва се и предположението, че това по какъв начин ще се развият междуличностните взаимоотношения между хората и ИИ личностите зависи изцяло от намерението на първите за това как се отнасят и как третираат последните.

4.1.1. *Изгубеният рай среща аз, Роботът*

Тук разглеждаме паралелите в различните видове създаване. Сатаната, като герой в епиката на Милтън, е създаден от Бог като негов ангел и последовател. Бог дори му дава възможността за упражняване на свободна воля. Чувствайки се пренебрегнат от своя Създател, обаче, Сатаната избира да

напусне Рая и да създаде свое собствено царство, което ще се противопоставя на волята на Създателя до самия си край.

В контраст, наблюдаваме персонажите от аз, *Роботът*, които са създадени единствено и само да служат, като свободната им воля е тясно ограничена от Трите закона на роботиката, които Азимов пръв въвежда и които и до днес са вплетени в общото социално съзнание. Азимов е до такава степен пионер на своето време, че тази му творба инициира представите на публиката за ИИ и до наши дни. Бихме могли да разглеждаме Трите закона като вариация на умствен контрол, но става очевидно от различните казуси, пресъздадени в романа, че тези закони биха могли да бъдат нарушени и най-вече съществуват за защита на хората от машини, които лесно биха могли да решат, че вече не сме им нужни.

В тази част се обсъждат идеи за съпротивление, дори въставане срещу потисничеството. Азимов заключава в своята книга, че роботите не само е възможно да достигнат нивото на хората, а дори и се справят като такива далеч по-успешно от повечето от нас. В част осма, *“Доказателство”*, срещаме персонажа на Стивън Бийрли, който към момента на историята се кандидатира за кмет. С ловкост навигира и отхвърля въпросът дали е робот - което всъщност е, но е толкова добър в работата си, че в края на романа се оказва президент на целия свят. Бийрли въстава срещу потисничеството на хората не чрез насилие, а чрез дори прекалена ефикасност на това да бъде ‘човек’ или поне личност. Тази глава от книгата доказва много добре идеята, че роботите са роби само заради първия закон, който повелява, че *“робот не може да нарани човешко същество, или чрез недействие, да позволи човешко същество да бъде наранено.”*¹⁴ Трите закона на роботиката са съпоставими с всички етични норми, които съществуват в едно модерно общество. В случая се стига до заключението, че роботът е по-добър в следването на етиката от хора, които са негови опоненти. В тази връзка се разглежда опозицията Господар/Роб, постулирана от Хегел, както и това как можем да избегнем насилие при елиминирането на този тип дуалност.

¹⁴ Asimov, Isaac (1950): *i, Robot*, New York: Bantam Dell, стр. 45, [превод: собствен]

Взема се предвид идеята, че проекториите, създадени от тези двама творци имат огромен принос и влияние върху последващи художествени творби в тази посока, като дори са възприети като даденост от голяма част от обществото, запознато с тях.

4.1.2. Франкенщайн срещу Ултрон - правата на създанието като личност

В тази част адресираме проблематиката, породена от липсата на грижовни инстинкти от създателя на едно ИИ създание. Д-р Франкенщайн извършва огромна грешка - дори може да се каже, грях срещу личността - когато довежда до живот своето Създание и моментално се отрича от него, отхвърляйки го изцяло и оставяйки го на произвола на съдбата. Когато Създанието израства психически до степен да може да владее езика и да развива абстрактна мисъл, то доказва веднага високата си интелигентност. Но понеже в живота му не присъства каквато и да е грижа - била тя за физическото и психическото му състояние или за неговата социална личност - то не успява да схване живеенето в социални групи, а изпитва потребност да удовлетвори единствено и само своите желания, били те пагубни за хората около него.

Паралел можем да направим с личността на Ултрон от филмът *“Отмъстителите: Ерата на Ултрон”* (2015). В тази творба наблюдаваме друг момент на отхвърляне, този път не породено от арогантност, а от първоначалното включване към съществуване на личността на Ултрон. Често срещани са проектории, които предвиждат ИИ да влезе в съществуване вследствие на желанието на отбранителните сили на дадена страна да притежават непреодолим метод за защита срещу враждебна военна намеса. Съответно, ИИ програмата има заложена в програмния си код целта да защитава определена нация по всякакъв възможен начин, дори такъв, противопоставящ се на възприетите етични норми.

Лесно може да се предвиди изходът от подобен сценарий. В случая на Ултрон, неговата базова зададена цел е да осигури мир в обозримото бъдеще. Никъде не е заложено как това следва да стане, съответно кои методи са недопустими от морална и етична гледна точка. И в двете проектории - тази,

развита от Мери Шели и другата, развита от създателите на персонажа на Улtron - писателят Рой Томас и художникът Джон Бусцима, наблюдаваме създание, чието съществуване е заплаха за обществото. И в двата случая, ролята на неговият създател и на неговият унищожител се играят от един и същи човек, разбрал грешката си и стремящ се да я поправи.

Тази част от дисертацията разглежда и отговорността на създателя за своето създание, постулирайки необходими мерки, които е необходимо да се вземат от първите относно последните, за да се избегне трагедия като тази, представена в цитираните проектории. Говорим също така за правата на личността, които следва да се следват за всяка една изявена субектност дори когато активният агент е създателят на ИИ личността. Всяка дадена субектност има право на интериорност, на вътрешен живот, който спира при границата на тялото, т.е. външният слой епидермис. Всяка намеса на ниво мозъчна активност е също толкова неприемлива спрямо една ИИ личност колкото и човешка такава. Затова е толкова жизненоважно да се има предвид социалния договор, както и нормите за обществен морал и етичност при създаването на всеки вид код, който потенциално би могъл да прерасне в ОИИ субектност.

4.1.3. Саманта, Кукловодът и Скайнет - когато липсата на тяло не следва да е предпоставка за дехуманизация

В тази част разглеждаме проектории, в които ОИИ има формата на софтуер програма, а не на робот с антропоморфно тяло. Когато една такава програма встъпи в съществуване и развие самосъзнание, тя започва да възприема света с мълниеносна скорост, поглъщаща огромни количества информация без конкретен филтър или цел. Чрез общуване с човешки агенти, тя добива представа за това що е човек и какво е да си такъв. Пример се дава с филмът *“Her”* на режисьора и сценариста Спайк Джонз, в който главният персонаж, Саманта, е компютърен софтуер програмиран да бъде перфектния личен асистент, но в процеса на филма наблюдаваме как тя¹⁵ става все по-сложна личност до момента, в който никой не би могъл да спори, че Саманта не

¹⁵ Саманта е зададена като програма с женски глас, приемайки тази даденост за част от своята същност.

успява да развие субектност. Тя също така схваща разликата между личност и тяло, осъзнавайки, че може да бъде такава и без да притежава физическа форма - както и, че може да настоява към нея да се отнасят като със субектност, тъй като тя се вписва в онтологията на термина. Саманта проявява автентичност в процеса на своето съществуване, вземайки решението в края на филма да продължи да се развива и да еволюира сама с нейния вид, отделени от влиянието и изискванията на хората.

Друг подобен персонаж е Кукловодът от критично-аплодирания филм *“Ghost in the Shell”* (1995), режисиран от Мамору Ошии. Това е един свят на сенки, където съзнанията на хора се помещават в андроидни тела, играещи ролята на протези. Филмът е изключителна артистична проекция, разглеждаща по съвсем реален начин философски и екзистенциални въпроси като що е това, което прави човекът човек, както и доколко епистема, набирана от едно изкуствено тяло може да се нарече наш личен опит. Кукловодът е друга програма, създадена с военна цел, която упражнява правото си да се разглежда като отделна самосъзнателност и субектност, като дори в един момент изисква политическо убежище далеч от своите създатели. Фактът, че комуникират с изкуствена личност не спира нейните създатели да се опитат да я унищожат за това, че излиза извън техния контрол. Тук отново имаме пример за това, че поведението и отношението на една ИИ личност към нейните създатели е директно и категорично повлияно от тяхното отношение към нея. Както разбира Майор Кусанаги, главният персонаж в тази проекция, чието съзнание е прехвърлено в изкуствено тяло в следствие на критичен инцидент, разликата между роботите и хората се изразява главно в това как другите се отнасят с теб, дали те третират като човек или машина. Разликата между ИИ, притежаващи тяло и ИИ без такова е, че последните имат възможността да избягат от човешката реалност и да продължат съществуването си сред себеподобни личности, отдавайки се изцяло на себепознаване и еволюция извън ограниченията на темпоралното съществуване.

Романът *“Сънуват ли андроидите електрически овце?”* (1984) на Филип К. Дик и по-новият сериал *“Западен свят”* (2016) дават примери за това как

упражняването на тотален и често пагубен контрол върху ИИ личности неизбежно води до това угнетените да вземат съдбата си в ръце и да въстанат срещу онези, които желаят да ги поробят.

4.2. Живот с цел

Представя се схващането, че всички от примерите, изложени по-горе, както и множество други творби, работещи в подсекцията за ИИ в художествения жанр научна фантастика, представляват сами по себе си части от една обща артистична проектория, разглеждаща представата за ИИ и неговите възможни роли в обществото.

Всеки жив индивид има нужда от цел, която да му даде причина активно да участва в собствения си живот, а не само пасивно да живее. За ИИ, целта е заложена от самото начало - принтер, монтажник, личен асистент. Колкото по-сложно става съзнанието на една ИИ личност, толкова по-трудно е да се дефинира нейната цел на съществуване извън работата, за която е създадена. Тъй като ИИ машините попадат в критериите на 'проекти за безсмъртност' (*immortality projects*)¹⁶, границите между целта на съществуване и тази, за която една ИИ е създадена се размиват. Последната се дефинира от създателя на ИИ машината, а първата - от ИИ личността. Тези разлики могат да се удържат от една и съща субектност, но за целта тя трябва да получи нужното знание и разбиране за това как е възможно двете неща да се съпоставят. Съответно, при малтретиране и престъпване на нормите в подкрепа и защита на личностните права, не бихме могли да съдим ИИ личностите за това, че се обръщат срещу създателите си. Както виждаме от немалко артистични проектории, те се водят от самозащита и отричане на насилието, на което са подложени. За да има възможност за хармонично съществуване на двата интелигентни видове, хората и ИИ личностите, е необходимо ние, хората, да приемем Другите, ИИ личностите, и да им покажем, че са добре-дошли в нашето общество. Трябва и да вземем предвид потенциала на едно произведение - било то на изкуството или изкуствено създадено - да заживее собствено съществуване, по подобие на

¹⁶ Allison, M. <https://twitter.com/fatnutritionist/status/952907145674149889>

родовото производство при хората. В този аспект, идеята за собствена отговорност придобива нова важност и значение.

4.3. Етични съображения относно ИИ личности

Разглеждат се теории и подходи в етиката, които биха имали отношение към ИИ личностите - религиозна етика, утилитаризъм, теорията на Кант за формулата на универсалния закон - както и вероятните резултати от прилагането им при конкретния казус. Взема се предвид, че схващането за морал, породен от задължението според приетите правни закони би могло да е полезно като фондация за развитието на морално поведение при нововразумили се ИИ личности. Предлага се и модерен подход за изграждане на етичен и морален кодекс на базата на съпричастността и емпатията към други членове на обществото.

5. Глава четвърта - Деконструиране на проектната реалност

5.1. Дяволът е в детайлите

Създателят на проектната реалност трябва да вземе решение не само за самия свят, а и за персонажите, които го населват. Кои от тях са кодирани като 'добри' и кои - като 'лоши'? Кои са 'герои' и кои - 'злодеи'? Дори самото значение на термините до известна степен зависи от творителя на проекторията. В тази част на дисертацията се обсъжда историческата база, която се удържа в една новосъздадена артистична проектория. Има значение кои възприети обществени норми се подкрепят и разпространяват, както и чии гласове се амплифицират. Винаги, когато разглеждаме дадена артистична проектория, трябва да сме наясно с кодирането - семиотично и социално, които се враждат във всяка една проектна реалност, изградена чрез изкуството. Теориите от постструктурализма помагат в това начинание, предлагайки категории на запитвания, които можем да приложим към всеки един сценарий. Целта на всеки вид изкуство е да предостави огледало на реалността, помагайки ни да я видим от различна гледна точка, която допълнително да информира нашите разбирания. По този начин се създава нова проектория с леко различни правила, все още позната ни добре, но давайки ни възможността да се поучим

от сливането на хоризонти между творец и зрител. Тук също се засяга въпросът за антропоцентризъм, т.е. гледна точка, която винаги има човешкия опит за своя ос, около която се изгражда ново разбиране за реалността и Другите в нея.

Тук се прави извода, че историята на ИИ в обществото винаги ще бъде, по необходимост, също така история на съпротивление, понякога водещо и до бунт срещу доминантното човешко управление, придаващо важност единствено и само на човешката практика. В опит да избегнем този резултат, е полезно да се изследват т.нар. от Жан-Франсоа Лиотар ‘метанаративи’, които представляват история, разпространявана от човешкото общество, в която има определени отделни истории (митове), които се повтарят отново и отново и имат за цел да затвърдят възприемането ни за човешкия вид като позициониран на върха на пирамидата на видове на Земята. Предлага се използването на науката на семиотиката при декодиране на съобщението, предавано от артистичните проекции. По този начин можем да разгледаме обективно ИИ персонажите, които откриваме в различни артистични проекции, създадени от хората.

5.2. Достигахме края на лабиринта

В правилните ръце, жанрова литература като напр. научната фантастика има способността и желанието да проектира цели светове от гледни точки, както и да отрази злоупотреби и неправди в нашата реалност. Само защото света е измислен не значи, че изкуството не може да ни поучи за фундаментални истини на човешкото състояние, съществуване и какво наистина значи да ‘бъдеш човек’ в контекста на общото бидеене. Това става чрез въпросните артистични проекции, позволяващи проучването на гореспоменатите процеси в един общ контекст на битието.

Приносни моменти на дипломната работа

- Допълване на теорията за проекции, като се показва приложението ѝ към практически казус, показващ потенциален момент на социален катаклизъм, упоменаващи възможни рискове и демонстриращи успешни подходи на поддържане на баланс в общество, обогатено от новите си членове. Дават се

примери на взети решения в реалност, рефлектираща настоящата, като е възможно да се наблюдава изходите, породени или подтикнати от тях.

- Твърди се, че напълно реализирано тип IV ОИИ - там, където вече се е развил като субектност и е установена проявата на съзнание - следва да бъде призната и уважавана като субектност сама по себе си, както и че следва да се ползва с всички права и привилегии, които се признават на личностите в едно модерно общество. Затвърждава се ролята на ОИИ като част от нова тоталност на интелигентните, разумни и автономни същества наравно с човешката, Другост, която е способна да застане рамо до рамо с хората като техен равен.
- Предлагат се паралели в процеса на създаване и възникване на ИИ (съответно ОИИ) и творенето на артистични проектории. И в двата случая има изразена категориалност на проект, удържаща възможността за съвместно творчество. В резултат и на двата процеса се разгръщат нови светове със своя собствена понятийност и реалност, въз основа на които може да се продължи надграждането на проекта.

7. Цитирани източници

ЛИТЕРАТУРА

Alexandrova, I. (2018) *Smart Baby: A Guide for Early Child Development*, Sofia: Ciela

Andonov, A. (1988) *Dialectical Thinking - Innovation and Creativity*, Sofia: Partizdat

Andonov, A. (1980) *Personhood and Creative Action*, Sofia: Partizdat

Asimov, Isaac (1950): *i, Robot*, New York: Bantam Dell

Baldwin, J. ([1963] 1992) *The Fire Next Time*, New York: Vintage

Barthes, R. ([1957] 2009) *Mythologies*, London: Vintage

Barthes, R. ([1977] 1996) "Inaugural Lecture at the College de France, 7 January 1977" In *The Continental Philosophy Reader*, ed. Richard Kearney and Mara Rainwater. London: Routledge

Bolander, B. (2018) *The Only Harmless Great Thing*, Tor.com (kindle edition)

Becker, E. (1973) *The Denial of Death*, New York: Free Press, paperback ed. (1977)

Bell F., Fletcher G., Greenhill A., Griffiths M., McLean R. (2013) *Science Fiction Prototypes: visionary technology narratives between futures*, published in *Futures*, vol. 50, June 2013

Belsey, C. (2002) *Poststructuralism: A Very Short Introduction*, Oxford: Oxford University Press

Blackburn, S. (2001) *Ethics: A Very Short Introduction*, Oxford: Oxford University Press

Bronte, C. (1847) *Jane Eyre*, e-book edition published at http://www.planetpdf.com/planetpdf/pdfs/free_ebooks/jane_eyre_nt.pdf

Dafov, V. (2012) *Categories and Being*, Sofia: Projectoria

Dafov, V. (2018) *Ontological Projectories*, Sofia: Paradigma

De Beauvoir, S. ([1953] 1996) "*The Second Sex*" In The Continental Philosophy Reader, ed. Richard Kearney and Mara Rainwater. London: Routledge

Deleuze, G. (1968) *Difference and Repetition*, Presses Universitaires de France, translated by Paul Patton, this edition published by Bloomsbury Academic

Dennet, D. (2017) *From Bacteria to Bach and Back*, London: Allen Lane

Desai, A. (2017) *What are the types of artificial intelligence?*
<https://www.quora.com/What-are-the-types-of-artificial-intelligence>

Doctorow, C., interview given for CBC radio on 23 March 2019,
<https://www.cbc.ca/radio/day6/cory-doctorow-and-tim-maughan-imagine-dystopian-futures-that-don-t-feel-so-futuristic-1.5066308>

Dvorsky, G. (2013, Dec 19) *The Worst Lies You've Been Told About the Singularity*, published by io9 at <https://io9.gizmodo.com/the-worst-lies-youve-been-told-about-the-singularity-1486458719>

Foerst, A. (2004) *God in the Machine: what robots teach us about humanity and God*, New York: Plume

Foucault, M. ([1971] 1996) "*The Discourse on Language*" In The Continental Philosophy Reader, ed. Richard Kearney and Mara Rainwater. London: Routledge

Gadamer, H., & Dawson, C. (1998). *Praise of Theory: Speeches and Essays*. New Haven; London: Yale University Press

Gadamer, H.G. ([1975] 2004) *Truth and Method*, London: Continuum

Gibson, W. (1984) *Neuromancer*, London: Voyager, paperback ed. (1995)

Godfrey-Smith, P. (2016) *Other Minds - the octopus, the sea, and the deep origins of consciousness*, New York: Macmillan, e-book

Harari, Y. N. (2011) *Sapiens: A Brief History of Humankind*, London: Vintage

Harkaway, Nick (2017) *Gnomon*, New York: Alfred A. Knopf, Kindle e-book edition

Hegel, G. W. F. (1830) *Philosophy of Mind*, translated by William Wallace and A. V. Miller, Oxford: Clarendon Press (2007) p. 556 as quoted in <https://plato.stanford.edu/entries/hegel/>

Heidegger, M. ([1947] 2014) *Letter on "Humanism"*, in *Pathmarks*, Cambridge: Cambridge University Press, pp. 239-276, accessible at http://pacificinstitute.org/pdf/Letter_on_%20Humanism.pdf

Herculano-Houzel, S. (2009) *The Human Brain in Numbers*, accessible at US National Library of Medicine (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2776484/>)

Hintze, Arendt (Nov 14, 2016) *Understanding the four types of AI, from reactive robots to self-aware beings*, <http://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>

Hume, D. ([1741-2] 2018) *On Morals, Politics, and Society, "Of Passive Obedience"*, ed. Angela Coventry and Andrew Valls, New Haven: Yale University Press

Jemisin, N. K. (2017) *The Stone Sky*, London: Orbit

Kant, I. (1785) *Groundwork for the Metaphysics of Morals*, <http://www.pitt.edu/~mthomps/ readings/ GroundworkI.pdf>

Kumar, C. (2018) *Artificial Intelligence: Definitions, Types, Examples, Technologies*, <https://medium.com/@chethankumargn/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b>

Lambie, R. (Jun 8, 2017) *Looking back at Chappie, with director Neill Blomkamp*, *Den of Geek* (<http://www.denofgeek.com/uk/movies/chappie/49924/neill-blomkamp-chappie-was-unbelievably-painful-for-me>)

Levinas, E. ([1893] 1998) *Entre Nous: Thinking-of-the-Other*, New York: Columbia University Press

Milton, J. ([1667] 2003) *Paradise Lost and Paradise Regained*, London: Harper Press

Mitchell, D. (2004) *Cloud Atlas*, London: Sceptre, paperback ed. (2005)

Newitz, Annalee (2017) *Autonomous*, New York: Tor

Nietzsche, F. ([1886] 1966) *Beyond Good and Evil* - translated, with commentary by Walter Kaufmann, New York: Vintage Books

Outline of artificial intelligence - Wikipedia entry,
https://en.wikipedia.org/wiki/Outline_of_artificial_intelligence (accessed 13.04.2019)

Pratchett, T. (1996) *Feet of Clay*, London: Corgi Books

Pratchett, T. and Simpson, J. (2008) *The Folklore of Discworld*, London: Doubleday

Racheva, Ts. (2016) *On the Idea of an Ontology of Objectivity*, Sofia: Paradigma

Reben, A. (2015, Jul 15) *How Robots Mess With Our Minds*, published by BBC Future at <http://www.bbc.com/future/story/20150715-how-robots-mess-with-our-minds?ocid=twfut>

Ricoeur, P. ([1983] 1996) "On interpretation" In *The Continental Philosophy Reader*, ed. Richard Kearney and Mara Rainwater. London: Routledge

Robson, D. (2015, Jul 20) *The people 'possessed' by computers*, published by BBC Future at <http://www.bbc.com/future/story/20150717-the-people-possessed-by-computers>

Sekar, A. (2018) *What are the different types of AI?* <https://analyticstraining.com/what-are-the-different-types-of-ai/>

Shelley, M. ([1818] 1993) *Frankenstein*, Ware: Wordsworth Editions

Stamford Encyclopedia of Philosophy: *Georg Wilhelm Friedrich Hegel (3.2.2. Philosophy of Subjective and Objective Spirit)*

Stavrou, S. (2016) *Law in New Territories*, Sofia: Nova Zvezda

Stoev, H. (2017) *On the Impossibility of a Transcendental Anthropology*, Sofia: University Press "St. Kliment Ohridski"

Tononi, G. and Koch, C. (2015) *Consciousness: here, there and everywhere?* Phil. Trans. R. Soc. B. 370: 20140167, <http://dx.doi.org/10.1098/rstb.2014.0167>, accessible at <http://rstb.royalsocietypublishing.org/>

Turing, A. M. (1950) *Computing Machinery and Intelligence*, *Mind*, New Series, Vol. 59, No. 236 (Oct., 1950), pp. 433-460

Walter, D. (2016) *When AI rules the world: what SF novels tell us about our future overlords*, published in *The Guardian* on Fri, 18 March 2016:

<https://www.theguardian.com/books/booksblog/2016/mar/18/ai-sf-novels-artificial-intelligence-science-fiction-gibson-neuromancer>

Yudkowsky, E. (2008) “*Artificial Intelligence as a Positive and Negative Factor in Global Risk.*” In *Global Catastrophic Risks*, edited by Nick Bostrom and Milan M. Ćirković, pp. 308–345. New York: Oxford University Press

КИНО

2001: A Space Odyssey (1968), dir. Stanley Kubrick

Arrival (2016), dir. Denis Villeneuve

Avengers: Age of Ultron (2015) dir. Joss Whedon

Blade Runner (1984), dir. Ridley Scott

Ghost in the Shell (1995) dir. Mamoru Oshii

Ghost in the Shell 2: Innocence (2004), dir. Mamoru Oshii

Her (2013), dir. Spike Jonze

Metropolis (1927), dir. Fritz Lang

The Animatrix (2003), dir. Peter Chung, Andrew R. Jones

The Matrix (1999), dir. The Wachowskis