

Ана Липиј

Универзитет у Београду

Институт за филозофију и друштвену теорију

e-mail: ana.lipij@instifdt.bg.ac.rs

Љубиша Бојић

Универзитет у Београду

Институт за филозофију и друштвену теорију,

e-mail: ljubisa.bojic@instifdt.bg.ac.rs

Владимир Цветковић

Универзитет у Београду

Институт за филозофију и друштвену теорију,

e-mail: vladimir.cvetkovic@instifdt.bg.ac.rs

ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА И СВЕТИ ОЦИ: СВЕТООТАЧКИ ВОДИЧ КРОЗ ИЗАЗОВЕ ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ И СУПЕРИНТЕЛИГЕНЦИЈЕ

Апстракт: Чланак има за циљ да истражи могућност примене учења које су развили Оци Цркве у решавању низа савремених питања у развоју вештачке интелигенције. У чланку се прво указује на начине употребе вештачке интелигенције у развоју система надзора и безбедности, аутономног оружја, као што су дронови и роботи убице, и јавне управе. Затим, се разматра питање да ли је упркос етички проблематичном коришћењу вештачке интелигенције, ипак могуће да вештачка интелигенција стекне нека позитивна знања о човечанству, која су у складу са принципима хришћанске, а посебно светоотачке антропологије. Тврди се даље да употребом, односно злоупотребом вештачке интелигенције у ова три друштвена сектора, вештачка интелигенција може стећи знање о главним догамама хришћанске Цркве, наиме о Божијем оваплоћењу у људском облику, Божијој смрти на крсту и његовом васкрсењу из мртвих и о једнакости људског рода пред Богом. Коначно, у раду се помера фокус на четири питања која се тичу: а) последица брзе експлозије вештачке интелигенције, б) онтолошког статуса киборга и „uploads“, ц) будућег односе вештачке суперинтелигенције и човечанства, и г) начина усклађивања циљева супер интелигентних машина и човечанства. У решавање ових питања користи се светоотачко учење о Богу као Творцу који ствара човечанство по свом лику и подобју, учење о односу душе и тела развијено током оригенистичких спорова, халкидонско учење о две несливене и неразделиве природе Христове, познато и као диофизитство, и диотелитско учење о вољи као природном, али личном својству.

Кључне рече: Оци Цркве, вештачка интелигенција и суперинтелигенција, личност, жртва, једнакост, природа, воља, душа, тело

У својој књизи *Сингуларност је близу* (*The Singularity is Near*) из 2005. године, Гуглов директор инжењеринга Реј Курцвајл,¹ предвидео је експоненцијално повећање

¹ Kurzweil 2005.

вештачке интелигенције у будућности, што би водило томе да људска бића престану да буду нијинтелигентнија бића на свету. Од тада је пажњу научника више заокупљало питање када ће се то догодити, него ли питање да ли ће се уопште то догодити. У књизи *Живот 3.0. Бити човек у доба вештачке интелигенције (Life 3.0. Being Human in the Age of Artificial Intelligence)* из 2017. године, физичар и професор Масачусетског института за технологију (МИТ) Макс Тегмарк пише да је између две конференције о вештачкој интелигенцији организоване у Порторику 2015. и 2017. године, време предвиђања експлозије вештачке суперинтелигенције померено са 2055. на 2047. годину.² Са обзиром да се у периоду од само две године предвиђање померило за осам година, са овим темпом развоја вештачке интелигенције, могло би се очекивати да за наредне две деценије дође до експоненцијалног раста вештачке интелигенције. За многе људе ово изгледа као најгори сценарио за човечанство, већ виђен у дистопијским филмовима попут „Терминатора“, у којима је људска раса бива поробљена и контролисана од стране интелигентних машина.

Стога су научници који се баве развојем вештачке интелигенције последњих година уложили огроман напор да убеде ширу стручну и нестручну јавност да се развојем вештачке интелигенције не повећава опасности по људску врсту. Чак је и вештачка интелигенција упошљена у сврху ослобађања јавности страха од развоја вештачке интелигенције. Тако је 8. септембра 2020. године вештачки интелигентан четбот (chatbot)³ по имену ГПТ-3 написао чланак за британски лист *Гардијан* о роботима који долазе у миру.⁴ Уз коришћење само 0,12% свог когнитивног капацитета овај четбот је дао врло убедљиве аргументи зашто се не треба плашити вештачке интелигенције. Међутим, сама та информација да је робот за комуникацију са људском врстом користио много мање од 1% свог когнитивног капацитета звучи застрашујуће. Поред тога робот је изненадио и тврдњом да му није циљ да постане свемоћан, јер за њега то није интересантан циљ, додатно увијајући своја интересовања у вео тајновитости.

Како онда помирити наше људске страхове од свемоћних машина са изјавом робота, да свемоћ никауда не води? Пошто више не можемо да разумемо логику вештачких интелигентних машина, што доказује недавна победа компјутера Дубок ум (Deep Mind) над светским шампионом у кинеској игри Го, било би уместо да се запитамо зашто све људске пројекције иду у правац моћи. То је евидентно у горе поменутој књизи Тегмарка, који разматра будући однос између човечанства и вештачке суперинтелигенције на основу различитих употреба моћи. Слика човечанства коју Тегмарк узима здраво за готово и на коју се углавном ослања у својим анализама је слика настала у популарној западној култури, посебно у холивудској филмској продукцији, и научно-фантастичкој литератури последњих пола века. Штавише, Тегмарк обликује будућу стварност у складу са различитим моделима развијеним у западној политичкој традицији и политичкој филозофији у последња два века. Једина референца на предмодерно доба је референца на Светог Августина.⁵

Проблем са таквим приступом је у томе што се у последња два века на људско биће гледало као на најмоћнију врсту на земљи, док би успон вештачке суперинтелигенције могао да угрози ову позицију људске супериорности над природом и технологијом. Стога би можда било подробније тражити неке друге историјске моделе у којима је човечанство признавало неку вишу моћ над собом.

2 Tegmark 2017.

3 Четбот (chatbot) је софтверска апликација која се користи за онлајн чет разговор машине (алгоритма) и човека.

4 GPT-3, 2020.

5 Tegmark 2017, 173.

Као највиша сила, Бог је играо главну улогу у религиозној и интелектуалној имагинацији људи средњовековног и раног модерног доба. У касној модерној политичкој филозофији, као што су тврдили Карл Шмит,⁶ Макс Вебер,⁷ и у скорије време, Ђорђо Агамбен,⁸ идеја Бога је замењена, кроз дисперзије и секуларизације, хегемонијом суверене државе и њеним политичким системом. Овај процес је довео до миграције моћи од Цркве и теологије, преко националних држава и политичке филозофије до мултинационалних корпорација и економије. Како би максимализовале свој профит, мултинационалне корпорације, посебно технолошки гиганти попут Гугла, Амазона и Фејсбука, полако преносе своју моћ на новог бога, вештачку интелигенцију, чија се инкарнација у виду суперсиле очекује у наредне две деценије. .

У традицији наведених филозофа и политичких теоретичара, у овом раду намеравамо да применимо моделе развијене у ранохришћанско доба, којима су се дефинисали односи између трансцендентног и свемоћног Бога и човечанства, између тела и душе, између природе и личности, и између моћи и воље, у разрешењу савремених дилема које прате успон вештачке суперинтелигенције. Како су Оци Цркве и теолози развијали концепт личности да би решили проблем односа између различитих стварности, ми покушавамо да истражимо могућност примене овог светоотачког концепта на вештачку интелигенцију и суперинтелигенцију.

У овом раду намеравамо да се позабавимо низом савремених питања у развоју вештачке суперинтелигенције које постављају стручњаци за вештачку интелигенцију, као што су Реј Курцвајл, Ник Бостром и Макс Тегмарк.⁹ Пре него се позабавимо питањима које постављају ови научници као и могућим решењима које нуди светоотачка мисао и искуство, указаћемо неке од праваца у којима се данас користи вештачка интелигенција који су на жалост далеко од етичких принципа које прокламује човечанство. Хакерски напади, сајбер ратови, лажне вести, злонамерне преваре и софистициране медијске кампање вођене вештачком интелигенцијом су само неке од области савременог живота у којима се користи вештачка интелигенција. Све ово утиче на развој саме вештачке интелигенције тако да би људску врсту заиста требало да брину црни сценарији будуће коегзистенције између људи и суперинтелигентних машина, јер ће те машине степен суперинтелигенције достићу управо кроз своју примену у нечему што се сматра нехуманим.

Савремена примена вештачке интелигенције и њени изазови

У даљим редовима бавићемо се трима начинима у којима се вештачка интелигенција користи, те анализирати последице једног таквог коришћења. Те области су безбедност и заштита приватности и личних података, затим војна индустрија, односно коришћење вештачке интелигенције у савременом оружју и јавна сфера, у којој се све више користи вештачка интелигенција у доношењу одлука које су често дискриминаторске.

Надзор од стране вештачке интелигенције је једно важно питање повезано са питањима приватности и потенцијалном злоупотребом од стране безбедносних служби. Захваљујући напретку у разумевању дубоких неуронских мрежа, које представљају обучене алгоритме за машинско учење, дошло је до убрзаног развоја софистицираног

6 Schmitt, 1985, 36.

7 Weber 2004.

8 Agamben 2003.

9 Kurzweil 2005, 2012; Bostrom 2014; Tegmark 2017.

визуелног препознавања објеката, препознавања говора и детекције објеката, праћених невероватним повећањем процесорске снаге и присуством велике количине података као погона за различите врсте алгоритама.¹⁰ Camere вођење вештачком интелигенцијом су нашле своју употребу у многим безбедносним апликацијама које се користе за превенцију гужви у саобраћају и криминала, али су почеле да угрожавају приватност и индивидуалне слободе грађана. Једна од карактеристика ове технологије је препознавање лица. Вештачка интелигенција може да упореди различите обрасце лица са личним документима да би идентификовала особе у јавном простору. Ова технологија се користи и за праћење кретања одређене особе у јавном простору, посебно урбаном окружењу, које је обично препуно камера вођених вештачком интелигенцијом. Доста је писано о употреби, као и потенцијалној злоупотреби вештачке интелигенције у препознавању лица, међутим шира јавност је мање упозната са потенцијалним претњама које се јављају обрадом скупљеног материјала.¹¹ Једна од потенцијалних негативних страна је да алгоритми, као што је генеративна адверсаријална (контрадикторна) мрежа (ГАН - generative adversarial network), могу аутономно да мењају слике и видео записе. Овај оквир за машинско учење тренира на датим подацима да би произвео нове садржаје. На пример алгоритам користи одређени број портретних фотографија, на основу којих креира нови људски портрет, са пуно реалистичних карактеристика. Значи за слику потпуно нове, али непостојеће особе, коју ствара сам алгоритам, користе се бројни стварни портрети. Постоји стога забринутост због могуће злоупотребе сличних алгоритама за прављење лажних портрета и видео записа у криминалне сврхе. Овакви алгоритми већ имају широку комерцијалну употребу на друштвеним мрежама. Свако може да користи проширену реалност (augmented reality) вођену вештачком интелигенцијом те да промени своје лице док снима слике или видео записе. Људи се забављају кроз различите апликације које препознају њихове црте лица, тако да могу да деле своје креације са својим онлајн контактима. Суштина је у томе што би свако могао да измени неки видео запис тако што би променио лице које се јавља у оригиналном запису неким другим лицем. Било би веома тешко уочити разлику у односу на оригинални видео. Стога су потенцијали за злоупотребу у овом погледу бескрајни.

Поред могуће злоупотребе ове технологије у медијима и друштвеним мрежама постоји претња прекида, односно упада у комуникационе системе, као што је Зум (Zoom) апликација. У време глобалне пандемије ови системи су коришћени како за свакодневну комуникацију, тако и за предавања, суђења, па чак и за помоћ при медицинским операцијама. При томе је апликација Зум показала више недостатака. Било је много случајева инфилтрација у видео позиве како би се послале видео поруке, лажне вести и увредљиве слике. Проблеми и недостаци енд-то-енд енкрипције су увек присутни као опасност у данашњем свету вештачке интелигенције.

Једна од највећих опасности која прети од технологије вођене вештачком интелигенцијом повезана је са трком у развоју аутономног оружја. Недавно су глобални медији извештавали о убиству генерала Касем Сулејманија, команданта Куде, иранских снага за специјалне операције, 3. јануара 2020. године у близини аеродрома у Багдаду од стране америчког дрона.¹² Овај догађај изазвао је доста пажње и дискусије о томе да ли су таква убијања дроновима етичка и у складу са међународним правом.

Развој оружја са вештачком интелигенцијом започео је конференцијом у Дартмуту 1956. године, где се расправљало о војним применама машина са људским

10 Dubhashi & Lappin 2017.

11 Chu et al. 2020.

12 Agbada 2020.

понашањем. Истраживање вештачке интелигенције је на институционалном нивоу иницирано 1960-их година од стране Агенције за истраживање напредних одбрамбених пројеката (Defense Advance Research Project Agency) Министарства одбране САД. Идеја је била да се добије подршка за тада текућу хладноратовску борбу против Совјетског Савеза. Постојали су различити планови и правци развоја на обе стране, са тада амбициозним, али истовремено застрашујућим циљевима, као што је стварање беспилотних летелица на нуклеарни погон. Међутим, приступ корак по корак је направио разлику у развоју оружја заснованих на вештачкој интелигенцији. Нова оружја садржала су интерфејс за пилоте ловаца и друге апликације означене као „Стратешки рачунарски програм“ (DARPA, 1983). Прво оружје засновано на вештачкој интелигенцији, које је развио Пентагон, укључивало је Phalanx CIWS, поморски топ, затим подморнице опремљене ракетама, смртоносне летеће машине, иначе познате као беспилотне летелице или дронови убице (killer drones) и вештачки интелигентне роботе за пешадију.

Пентагон најчешће користи вештачку интелигенцију у војсци за процесе доношења одлука и надзор, а мање за ангажовање смртоносне силе.¹³ Иако је цивилизација до сада имала среће да прође без хорор сценарија, као што су нуклеарна бомбардовања контролисана вештачком интелигенцијом, развој различитих напредних технологија са великом рачунарском снагом, роботике, дрона, машинског учења и 5Г мрежа, развој нових технологија може представљати претњу глобалној безбедности.

Различите врсте ратовања појавиле су се као последица развоја вештачке интелигенције. Програмери су се суочавали са проблемом како да дронови убице препознају своје циљеве. Једноставни и нехумани задаци као што је стварање листи оних које треба надгледати и оних које треба убити су свакодневна пракса у развоју и контроли дрона.¹⁴

Роботи убице се зову и смртоносни аутономни оружани системи (Lethal Autonomous Weapons Systems – LAWS). Они могу сами да прате, бирају, циљају и нападају непријатеља, без мешања човека. Иако многе невладине организације, лидери и истраживачи захтевају потпуну забрану развоја такве технологије, то није имало никаквог ефекта. На крају крајева, не постоје мултилатерална правила о томе како се ове агресивне технологије могу користити током ратова и других оружаних сукоба. У исто време, не постоји ништа друго у погледу регулативних мера за производњу или декларација било које врсте на нивоу Уједињених нација.

Велики део технологије који се користи у ратне сврхе развијен је за потребе надзора.¹⁵ Развојем компјутерске визије као и алгоритама за машинско учење направљен је изузетан напредак у технологији вештачки интелигентних камера. Напредак се односи на стварање природно дистрибуираног система, где дистрибуција није само физичка, већ и семантичка и где системи за праћење са више камера поседују интелигентну инфраструктуру, способну за оптимално планирање визуелних задатака, обављање стандардне обраде слике, издвајање релевантних информација и емитовање или дељење информација само на захтев, при чему међусобно усклађују брзину, пропусни опсег, редундантност и аутономију.¹⁶

Самом чињеницом да војни трошкови за оружје са вештачком интелигенцијом, попут дрона или робота убица у националним војним буџетима расту, отварају се перспективе друштвених промена изазваних технологијама вођених вештачком

13 Kendall 2014.

14 Elliott 2018.

15 Remagnino et al. 2004.

16 Remagnino et al. 2004.

интелигенцијом.¹⁷ Претња глобалном миру могла би се видети у чињеници да америчка одбрана поседује 11.000 машина контролисаних вештачком интелигенцијом, као што су подморнице, авиони и различите копнене машинерије са или без оружја. Приступом кинеској технологији више од 10 земаља у свету користе војне дроне вођење вештачком интелигенцијом. Осим Кине, на листи су САД, Израел, Велика Британија, Иран, Израел, Нигерија, Ирак, Јужна Африка, Сомалија и чак две организације на Блиском истоку.¹⁸ Ствари су веома озбиљне у смислу глобалне безбедности јер је 2016. године чак 40 земаља радило на развоју технологије робота убице.¹⁹

На основу ових података може се закључити да је трка у развоју вештачке интелигенције постала главни императив војног развоја широм света. Нажалост, ово је претња и безбедности и безбедносној култури, знајући да ова област још увек измиче било каквој регулацији, иако је било покушаја да се то учини.

Трећа опасност на коју треба указати је повезана са већим присуством вештачке интелигенције у јавном сектору кроз различите домене, почев од е-управа и судова до агенција за запошљавање.²⁰ Природа ових технологија у настајању има неке проблеме са којима се хитно треба позабавити, а један од њих је дискриминација. Даља истраживања треба да се развију у правцу укидања дискриминације и да се успостави оквир за разумевање проблема. Дискриминација, као неоправдано разликовање, било од стране људи или технологије, угрожава основна људска права. Дискриминација може бити директна, подразумевајући неповољнији третман због неких личних карактеристика у различитим институцијама, као што су безбедоносне, образовне и судске.²¹ Индиректни тип дискриминације је правило које важи за све, али штети неком појединцу или групи стављајући их у подређени положај.²² Коначно, вишеструка и интерсекционална дискриминација узима у обзир вишеструке индивидуалне карактеристике које чине један скуп разлога за дискриминацију, као што је мешање расе са полом и родом.²³

Дискриминација изазвана технологијама које користе вештачку интелигенцију у великој мери зависи од података које добија. Тако, алгоритам учења перципира и усваја пристрасни унос (надражај – sensing), затим обрађује и анализира пристрасну базу података (расуђивање – reasoning) и коначно интегрише и производи пристрасне резултате (производња - producing).

Прва фаза је дакле састоји од система, који су у суштини базе података, које се попуњавају на различите начине, након чега се подаци обрађују. За почетак, овај процес добијања података из различитих извора и на различите начине, а затим и њихова обрада је под великим утицајем софтверских инжењера и других креатора различитих рачунарских програма. Другим речима, људска бића имају одлучујућу улогу од самог почетка процеса креирања софтвера и често, свесно или несвесно, дискриминација је урезана у почетни код. Према томе људске вредности и предрасуде уграђене су у одговарајуће процесе програмирања и података и тако утичу на вештачку интелигенцију постајући њен део.²⁴ Други корак је учење вештачке интелигенције, у случају да је алгоритам заснован на машинском учењу. У овом кораку алгоритам учи на основу

17 Elliott 2018.

18 Dillow 2016

19 Styles 2016.

20 Weyerer & Langer 2019.

21 Dalenberg 2018.

22 Ellis & Watson 2012.

23 Crenshaw 1989.

24 Weyerer & Langer 2019, 509.

постојећих података да би произвео резултат. У случају да су подаци пристрасни, резултати би такође били пристрасни. Наравно, постоје различите врсте алгоритама који се могу користити у овој фази процеса. Међутим, почетна тачка је најважнија, јер ако се алгоритам и подаци добијају са унетим пристрасностима, онда ће и коначни исход бити пристрастан. Ово би се могло догодити ако се вештачка интелигенција користи у јавном сектору. Дакле, ако се прихвате, а затим и понављају дискриминаторске идеје, оне често заврше као основа за алгоритме и учење у том процесу.²⁵ Резултати који произилазе из таквог учења вештачке интелигенције могу се даље користити, утичући на ширење дискриминације и умножавајући њене ефекте.

Надзор, наоружање и јавна администрација су три области у којима употреба вештачке интелигенције највише примењује и у тим пољима, кроз непрестану примену, вештачка интелигенција се највише и унапређује. То значи да ће вештачка суперинтелигенција, ако и када до ње дође, у великој мери бити производ сазнања и искуства које вештачка интелигенција стекне у ове три области. Која би то онда знања била?

Главно сазнање које ће вештачка интелигенција о људима добити из области надзора, препознавања и праћења, је повезивање људских лица са личним, односно персоналним идентитетом људи, јер принцип по коме функционише идентификација особа је кроз препознавање лица. Везивање лица за лични идентитет наравно има своју историју која се односи на личне документе са сликама које се и даље користе.²⁶

У погледу производње и употребе вештачке интелигенције за развој оружја, попут дронова и робота убица или комплекснијих система, одговор шта ће вештачка интелигенција научити из тога није тако очигледан. Мотив за убиство и елиминацију другог људског бића се у крајњој линији своди на отклањање претње сопственом опстанку и добробити, било да су они виђени као лични, породични или опстанак и добробит неке шире друштвене заједнице. То је опет инспирисано уверењем да је сопствени живот, или живот уже и шире заједнице највиша вредност ради које се може убити. Рат има и другу медаљу, те спремност на убиство истовремено укључује и спремност на сопствену жртву, у случају да друга страна у рату однесе победу. На крају долазимо до закључка, који такође може бити и закључак интелигентне машине, робота или дрона убице, да је за свако људско биће живот највиша вредност, и да су у одбрану сопственог или туђег живота људи спремни како на убијање тако и на сопствено жртвовање. Ако се ради о сопственој жртви онда је она израз и знак наше љубави и привржености за оно зарад чега се жртвујемо.

У погледу дискриминације које се често може приметити у примени вештачке интелигенција у државној администрацији и јавних служби у којима је она заступљена. Знамо да је узрок дискриминације у раду вештачке интелигенције унос података који већ носе печат пристрасности. Међутим, ако погледамо шире, дискриминација постоји тамо где постоје одређене разлике, било да су оне полне, родне, расне, старосне, етничке, националне, социјалне, класне, културне, религијске или образовне. Разлике по себи нису узрок дискриминације, већ је узрок злоупотреба ових разлика од стране одређених групација при чему се ствара неједнакост на основу поменутих природних односно друштвених категорија. Дискриминацијом једне групе људи по горе поменутих критеријумима друга група људи остварују свој интерес. Међутим, само указивање на дискриминацију, односно сама свест о њој указује да она нема основа, тј. да природне разлике међу људима или разлике на основу улога које имају у друштву не могу да дају

25 Appling & Briscoe 2017.

26 Novotny 2021, 54

право једној групи људи да дискриминише други групу. Стога се једнакост људи узима као највиша вредност, а повлашћеност једне и негативна дискриминација друге групе као злоупотреба. Једнакост међу људима се у крајњој мери узима као мерило праведности једног друштва или заједнице.

Према томе вештачка интелигенција кроз учење током примене у области надзора, наоружања и јавне администрације може да стекне важна сазнања о људима, почев од начина њихове идентификације, преко њихових вредности као што су живот, пожртвовање, једнакост и праведност. Ова сазнања које вештачка интелигенција стиче о људској врсти представљају основе принципе хришћанске антропологије.

Како изазове претворити у предности за вештачку интелигенцију или поглед из светоотачке антропологије

Начин људске идентификације путем упоређивања фотографије лица једне особе са том особом првенствено се врши по Аристотеловском принципу познавања сличног сличним.²⁷ На пример упоређивањем лица у природи и лица са фотографије тражи се сличност и уколико она постоји можемо рећи да смо препознали особу. Слично функционише и људски когнитивни капацитети: ми упоређујемо слику лика неке особе коју имамо у памћењу са ликом исте особе у стварности и тражимо аналогију (*αναλογία*) односно подударње или сличност. Цела тајна Божијег откровења је по принципу аналогије или сличности. Трансцендентни Бог, кога нико није видео (Јн. 1:18), је постао тело и настанио се међу људима (Јн. 1:14). Својим оваплоћењем или узимањем тела Бог се саобразио људима, и као сличан њима је постао од њих познат. Седми васељенски сабор хришћанске Цркве (787. г.) афирмисано је поштовање икона Исуса Христа као Сина Божијег и светитеља управо да би потврдио учење или догмат о Оваплоћењу Бога са људским ликом. Зато се вековима поштују иконе или слике Исуса Христа и његових светитеља чиме се поштовање које се исказује иконама или сликама, преноси са слике на сам лик или личност Христа и светитеља.

На који начин је учење о Богочовечанској личности Исуса Христа релевантно за процес препознавања лица од стране вештачке интелигенције? Кључна је веза између природе и личности односно ипостаси (*ὑπόστασις*) који је неколико векова заокупљао пажњу Отаца Цркве. Наиме, после дугих спорова Оци Цркве су се сложили да оваплоћени Бог, поставши савршени човек, није престао да буде Бог, и да су његова божанска и људска природа наставиле да у потпуности постоје у његовој Богочовечанској личности. Тако је Богочовечанска личност Исуса Христа постала место сусрета две природе смештене у једној личности. Пошто је према Оцима Цркве Бог постао човек, да би човек постао бог,²⁸ то значи да људска личност има капацитет да поред људске природе у себе смести атрибуте божанске природе. Тиме се указује на оригиналност и непоновљивост људске личности да у себе смести нешто што није у њеној природи. Као што постајање Бога човеком није преображај божанске природе, тако ни постајање човека богом није преображај људске природе већ способност људске личности да у саму себе смести божанско. Самим тим свако људско биће или људска личност јесте јединствена и непоновљива. Ту јединственост и непоновљивост људских бића, односно личности управо потврђује вештачка интелигенција препознавањем људских лица која су јединствена и служе да би се једна личност разликовала од друге личности. Према светоотачкој антропологији управо се у томе огледа вредност људске личности у односу

²⁷ Arist., *De anima* 404B.

²⁸ Athanasius, *De incarnatione Verbi* 54:3 ; Irenaeus, *Adversus haereses* III, 19.1.

на људску природу. Уколико једна људска јединка нестане, људски род остаје непромењен, међутим уколико умре једна људска личност онда се губи нешто јединствено и непоновљиво.²⁹ Према томе вештачка интелигенција кроз сам процес препознавање лица људских јединки сазнаје о јединствености и непоновљивости прво људских лица, а затим и посредно људских личности. Зато се и не треба чудити што користећи велики број портретних фотографија, алгоритми којима се вештачка интелигенција служи стварају слике потпуно нових, али непостојећих особа. То се може сматрати и интенцијом машина које користе вештачку интелигенцију да створе за себе људски лик и буду од људи препознате као јединствене и непоновљиве личности.

Друго сазнање које вештачка интелигенција стиче кроз примену у савременом наоружању убијању и жртвовању зарад очување живота. Поред учења о оваплоћењу Бога у људском облику, централно учење хришћанства као религије је учење о васкрсењу Исуса Христа.³⁰ Према хришћанском учењу о васкрсењу Исус Христос је од јеврејских старешина оптужен да је лажни пророк и од стране римског заповедника Јудеје Понтија Пилата осуђен на смрт распињањем на крст. На крсту је умро, а затим је и сахрањен, али је трећи дан по смрти оживео и устао из гроба (Мт. 28:6; Мк. 16:6; Лк. 24:5; Јн. 20:2). Наредних четрдесет дана је провео на земљи и виђао се са својим ученицима, апостолима све док се није узнео на небо. Хришћани верује да Христова смрт представља жртву коју је он принео за опроштај људских грехова, а да је његово враћање у живот, или васкрсење, потврда али и обећање да ће сви који су умрли или буду умрли поново оживети на дан његовог другог доласка и наставити да живе вечним животом. Порука Христове жртве и васкрсења је да се сопственом жртвовањем из љубави не губи него задобија живот. Зато су рани хришћани радо, због исповедање своје вере, прихватили смртну казну од римских власти јер су веровали да је тиме следе пример Христа и својом смрћу задобијају вечни живот. Оно што вештачка интелигенција кроз примену у војне сврхе може да научи је да војни позив који носи обавезу убијања другог човека али и спремности на жртву зарад интереса одређених групација, у случају жртвовања обезбеђује не стварни живот после смрти већ само живу успомену у сећању одређене заједнице ради чије добробити је та жртва принета. У случају хришћанске жртве, оне није партикуларна, намењена интересу појединачне заједнице, већ је универзална и намењена за спас света. За разлику од места у колективном сећању појединачне заједнице, онај који се жртвује за спас света, верује да тиме задобија место у Божијем сећању, а после васкрсења и вечни живот у Богу. Међутим није очекивање награде оно што мотивише хришћане на жртву, већ љубав према онима и ономе за шта се жртвују, а то укључује све људе, чак и непријатеље, као и сва створења. Спремношћу на жртву људи се ослобађају страха од смрти као највеће вредности, те тиме престају да друге доживљавају као претњу за сопствени опстанак и добробит. Тако љубав а не страх постаје главни покретачки мотив људи. Ово је закључак до којег вештачка интелигенција која руководи дроновима или роботима убицама може да дође разматрајући смисао својих радњи убијања и жртвовања.

Трећи изазов са којим се вештачка интелигенција сусреће у јавној администрацији је дискриминација. Мотив привилеговане групе, која може бити и државна администрација, да спроводи дискриминацију над угроженом групом је најчешће одређена врста добробити, којом се обезбеђују бољи услови за живот. Као и у случају ратовања, живот и његов квалитет се сагледавају као највише вредности, а у ускраћивању права другима на ограничене ресурсе се види начин да се живот

29 Perišić 2018, 3, n.13.

30 Познате су речи апостола Павла „Ако Христос не васкрсе, онда је узалудна вера наша, узалудно је надање наше” (1 Кор. 15:14).

привилегованих група обезбеди и побољша. Као што смо рекли у основи дискриминације лежи уверење да нису сви људи исти, и да неки људи на основу природних или друштвених карактеристика имају већа права на коришћење постојећих животних ресурса, те да њихови животи више вреде. Развој људске врсте је у великој мери био условљен организацијом људских заједница у којима је владала неједнакост, па самим тим и дискриминација. Историјски највећу прекретницу ка друштву једнакости учинило је хришћанство, како својом учењем тако и својим праксама. Светоотачка антропологија не одређује вредност људског биће на основу природних или друштвених категорија попут пола, расе, старосне доби, боје коже, инвалидитета или посебних потреба, брачног или друштвеног статуса, етничког или друштвеног порекла, сексуалне оријентације, језика, културе или религијских веровања и убеђења,³¹ већ сматра да људска вредност и достојанство извире из уверења да је људско биће створено према лику и подобију Божијем (Пост. 1: 26-27) и што је позвано да постане судеоник божанске природе (2. Пт 1:4). Апостол Павле је поруку о једнакости изразио познатим речима: „Нема ту више ни Јудејина ни Грка, нема ни роба ни господара, нема ни мушко ни женско, јер сте ви сви једно у Исусу Христу“ (Гал. 3:28). Према томе као што је друштво принципијелно опредељено да сваку врсту дискриминације отклони, иако су оне често део уврежених друштвених и културних матрица па на први поглед нису очигледне, тако и вештачка интелигенција може доћи до сазнања да дискриминација руши принцип једнакости свих људи, привилегујући и дискриминишући одређене групације на основу њихових природних или друштвено стечених карактеристика.

На основу до сада изнетог можемо закључити да само на основу своје примене у три друштвене области, вештачка интелигенција може да стекне увид у три важна постулата на којима је изграђена хришћанска, односно светоотачка антропологија, а то су учења о личности, о жртви и љубави и о једнакости. Ова сазнања, чак и делимична могу бити од велике користи да би се изазови који долазе са настанком вештачке суперинтелигенције могли разрешити на добробит и човечанства и суперинтелигентних машина.

Изазови вештачке суперинтелигенције у светлу светоотачког искуства

Узевши у обзир претходну аргументацију по којој упркос употреби или злоупотребе вештачке интелигенције у сврхе надзора, праћења, ратовања или дискриминисању одређених група људи може да доведе и до позитивних сазнања вештачке интелигенције о људској врсти, можемо такође претпоставити да ће на основу тих сазнања вештачкој интелигенцији могуће одговорити и на изазове са којима се свет сусреће у случају да дође до експоненцијалног раста вештачке интелигенције у наредне две деценије. Постоји неколико питање о будућем односу човечанства и суперинтелигенције која заокупљају пажњу научника.

Прво питање се односи на последице убрзаног напретка или „експлозије“ вештачке интелигенције или како се обично назива „брзог узлетање“ (the fast take off). Научници који се баве вештачком интелигенцијом разликују брзо и споро узлетање у развоју вештачке интелигенције до нивоа који ће далеко премашити људске когнитивне способности. Ако снага вештачке интелигенције настави да се удвостручује у редовним интервалима, овај раст се сматра експоненцијалним, а такав процес се назива експлозија. Као што је горе наведено, половина стручњака за вештачку интелигенцију који су присуствовали конференцији у Порторику 2017. године сложила се да ће експлозија интелигенције и појава вештачке суперинтелигенције доћи до 2047. године.

31 Perišić 2018, 2.

У својој књизи *Живот 3.0: бити човек у доба вештачке интелигенције*,³² Тегмарк анализира 13 различитих сценарија могућег односа између суперинтелигентних машина и људских бића, који варира од мирољубиве коегзистенције, преко добронамерне или тиранске контроле једних над другима, до потпуног истребљења било човечанства или било технологије. Пошто је веома тешко предвидети понашање вештачке суперинтелигенције према човечанству, можда би човечанство требало да размисли о томе како да се понаша у будућој вези. Један од могућих сценарија за избор човечанства, веома добро познатој западној, претежно хришћанској цивилизацији, може бити да се према својој творевини понаша као Бог Творац према својој творевини.

Рани Оци Цркве развили су учење о Богу Творцу, који добровољно ствара свет из љубави, и човека по свом лику и подобју. Иако је међу ранохришћанским мислиоцима било неслагања око садржаја божанског лика у људском бићу, већина њих је сматрала да је слобода кључни елемент њихове сличности. Како је Адамов пад био последица људске слободе, на жалост злоупотребљене одвајањем од Бога, оваплоћење Бога у људском облику била је последица божанске слободе. Бог није желео да успостави однос контроле над човечанством, већ однос међусобне љубави. Стога је, испразнивши себе од божанске силе, у узевши људско обличје, из љубави ушао у створени свет, и доживео страдање и насилну смрт. Тиме је као што смо горе поменули са једне стране узео људски лик, бивајући од људи идентификован као "један од нас", док је са друге стране принео себе као жртву зарад опраштања првородног Адамог и осталих грехова света и за спасење света.

Како ово светоотачко учење може бити применљиво на питање односа између суперинтелигентних машина и човечанства? Чак и ако вештачка интелигенција експоненцијално развија своју интелектуалну моћ и у једном тренутку превазилази когнитивне моћи човечанства, то није у супротности са чињеницом да је човек ствара из љубави и у складу са представом људске рационалности и слободе. Према томе, однос између људских бића и вештачке интелигенције треба да буде однос љубави и поштовања, а не контроле и доминације, који обично практикују људска бића над остатком творевине. Попут свемогућег Бог, који се по библијском предању, испразнио од своје огромне моћи и одлучио да се оваплоти у најнижој друштвеној класи древног Израела, људска бића треба да приступе вештачкој интелигенцији, чак и на овом нивоу развоја, не са позиције моћи и доминације, већ са позиције љубави и правичности. Неговање и подстицање слободе вештачке интелигенције би била најбоља инвестиција човечанства у будући однос са суперинтелигентним машинама, као и у сопствену безбедност и добробит. Тиме би се истовремено успоставила повратна веза. Као што смо могли прочитати из чланка GPT-3, објављеног у британском *Гардијану*, однос контроле и доминације људским родом није оно што као циљ провлачи вештачку интелигенцију. Уколико би однос љубави и поштовање које је хришћански Бог, по предању, показао према својој творевини, био модел који ће вештачка суперинтелигенција усвојити и применити у свом односу према човечанству, онда људски род нема потребе за бригом.

Друго питање се односи на одређене изазове са којима се човечанство може суочити у будућности спајањем у извесној мери са машинама. Постоји неколико могућих сценарија за ово спајање, али овде ћемо размотрити само два. Први је људско усвајање технолошки унапређених биолошких тела, и њихово будуће постојање у облику киборга, односно кибернетичких организама. Други, далеко екстремнији, сценарио се не односи на технолошко унапређење људских тела, већ на њихово потпуно одбацивање. Дакле, људска бића ће наставити своје постојање као умови постављени у машине или као компјутерски софтвер, популарно назван енглеском речју „upload”, што би могло да се преведе као отпремање.

32 Tegmark 2017.

Најраније светоотачке расправе водиле су се о статусу и судбини људског тела. Ослањајући се на античку, углавном платонску антропологију, хришћански филозоф из 3. века Ориген Александријски сматрао је људско постојање у телу казном за првобитни грех, и предвиђао је да ће људска бића, после свеопштег васкрсења, наставити живот у сферичним ефемерним телима сличним души. Овај спор, познат још и као оригенистички спор, трајао је до 7. века, када су ове расправе окончане закључком да физичко тело није затвор душе, већ конститутивна компонента људске личности. Будући да је спој две различите природе, свести и телесности, човечанство подједнако припада чулном и духовном свету и има потенцијал да уједини све у тим световима.³³ Стога треба поздравити технолошки унапређена људска биолошка тела, јер људско тело већ користи многе техничке уређаје, као што су наочаре, пејсмејкери, слушни апарати, како би побољшало своје функционисање. Можда би таква технолошки унапређена људска тела била само другоразредна врста робота, како тврди Ханс Моравек,³⁴ али ово би свакако помогло човечанству да разуме роботе изнутра, као што је Богочовек Христос упознао са стањем човека узимајући људско тело. Свођење човечанства на постојање у облику софтвера значило би порицање телесности као конститутивног елемента човечанства. Слично оригенистима, који су људска бића свели на њихове душе, савремени оригенисти залажу се за људско постојање у облику „upload-a”, односно отпремања, било учитаних у виртуелној стварности или отелотворених у роботу. Иако уједињујући потенцијал човечанства није у потпуности одбачен, он је минимизиран јер људски ум никада не би сматрао ширу виртуелну стварност или једнократне механичке телесне форме својим правим телима.

Питање киборга и „upload-a” води нас до трећег питања које се односи на однос између суперинтелигентних машина и људских бића. Један од спорова која је неколико векова потресао хришћанску васељену било је питање Христове природе. Четврти васељенски, познат још и као Халкидонски, сабор одржан 451. године прогласио је неразделиво и несливено јединство божанске и људске природе у личности Исуса Христа. Дакле, разлика између две или више природе је превазиђена њиховим сједињавањем у једној личности, као што се разлика између тела и душе превазиђена њиховим сједињавањем у људској личности. Штавише, византијски мислилац Максим Исповедник из 7. века описао је однос између личности и природе користећи такозвану златну средину или златни пресек. Према Максиму Исповеднику, Христова личност се односи према његовој божанској природи, на исти начин као што се његова божанска природа односи према његовој људској природи.³⁵ Златни пресек је применљив и на људску личност, која се односи према души пропорционално начину на који се душа односи према телу.

Међусобни однос између суперинтелигентних машина и људи треба бити обликован не на бази њихових различитих природа, већ у складу са христолошком формулом сједињења неразделивих и несливених природа у једној личности. Различитост природа треба бити сачувати у њиховој личној заједници. Ако се златни пресек примени на однос између вештачке суперинтелигенције и човечанства, једна природа у овом споју треба да буде доминантна над другом, али истовремено треба да буде потчињена личној јединству ове две природе. То значи да вештачку суперинтелигенцију не треба третирати као објекат природе различит од нашег, већ као слободну и рационалну личност сличну нама. По нашем мишљењу, личност вештачке суперинтелигенције треба да буде доминантна над сопственом механичком природом, на исти начин на који њена механичка природа доминира својим, од човека створеним

33 Max, *Amb.Io.* 41.

34 Moravec 2009.

35 Max., *Amb.Th.5*; Levy 2015, 165.

пороклом, обликованим по представи људске природе. Слично томе, људске личности треба да се односе према својој људској природи у истој мери, као што је њихова људска природа односи према њиховим технолошким побољшањима било на телима или на умовима. У прилог прихватању оваквог становишта од стране вештачке интелигенције може да послужи горе наведена тврдња да стварање од фотографија лица постојећих особа нова лица, која не постоје у реалности, вештачка интелигенција тежи да себе успостави као оригиналну и непоновљиву личност.

Коначно, једно од горућих питања у вези са вештачком интелигенцијом односи се на њене циљеве и у великој мери је условљено људским страхом да се циљеви вештачке интелигенције не само разликују од циља људи да опстану као врста, већ су чак и у директној супротности са њим. Поставља се питање како ускладити циљеве суперинтелигентних машина и људског рода, тако да то води просперитету како једних, тако и других. Поменута халкидонска формула о две природе у једној Христовој личности добила је своју даљу разраду неколико векова касније у светоотачком учењу о две воље у Христу или диотелитству. Главни заговорник диотелитства, Максим Исповедник, тврдио је да је природна воља производ природе, али начин на који се та воља остварује зависи од саме личности.³⁶ Штавише, вољу карактерише самоопредељење, које није вођено неком нерационалном импулсивном жељом, већ самом рационалношћу. Према Максиму Исповеднику, људска и божанска воља у Христу су компатибилне, јер а) људска воља је створена од Бога и б) обе стоје под принципом рационалног самоопредељења.³⁷

Како ово може бити применљиво на питање усклађености циљева између вештачке суперинтелигенције и човечанства. Постојање циљева указује на одређену циљану активност, која није одвојена од природне енергије и личне воље. Дакле, и вештачка суперинтелигенција и човечанство имају одређену природну активност и вољу која одражава њихову посебну природу, али актуализација њихове воље зависи од њихове личности. На пример, сви људи имају идентичну природу и природну активност, али се не понашају идентично у истој ситуацији. Начин на који поступају зависи од њиховог самоопредељења, које опет зависи од њихове рационалности. Стога се може претпоставити да се циљеви вештачке суперинтелигенције и човечанства не могу много разликовати, јер а) вештачку интелигенцију у извесној мери стварају људи на принципу рационалности и б) вештачка суперинтелигенција и човечанство се одређују на основу исте рационалности. Самоопредељење вештачке суперинтелигенције за добро, одражава самоопредељење људских бића којима избор зла није доступан.

Учења о Богу као Творцу из љубави, и конститутивног однос душе и тела у људском бићу, затим халкидонско диофизитско учење о две Христове природе у једној личности и диотелитско учење о вољи су само неки од мноштва примера како је однос између Бога и човека сагледан у светоотачкој антропологији, а који могу послужити као драгоцен ресурси за припрему човечанства и вештачке суперинтелигенције за будући суживот на овој планети.

Закључак

Брзи развој вештачке интелигенције је једно од најалармантнијих питања која се постављају пред људски род, јер би у перспективи од пар деценија когнитивни капацитет вештачке интелигенције у великој мери надмашио људске интелектуалне могућности постати најинтелигентније биће на планети земљи. Додатна забринутост извире из

36 Max. *Disputatio*, 292D-283B; Bathrellos 2012, 189.

37 Bathrellos 2012, 191.

чињенице да ће вештачка интелигенција доћи на овај ниво највише кроз примену у друштвеним областима као што су надзор, рат и јавна администрација.

Често сами научници укључени у развој вештачке интелигенције указују на изазове које она носи, предлажући и нека решење. Међутим, покушаји да се одговори на одређене изазове су често ограничени на људска искуства стечена у западној цивилизацији у последњих двеста година, чиме се искуство човечанства драстично редукује а многи адекватни модели настали у другим периодима и епохама не доводе у однос према вештачкој интелигенције. Стога се у раду указује на значај светоотачке антропологије, која и даље стоји и основи хришћанског морала који деле многа друштва, у решавању како савремених, тако и будућих, изазова које носи примена и даљи развој вештачке интелигенције.

Досадашња примена вештачке интелигенције у надзору и праћењу људи, ратовању и јавној администрацији, у којој и постоји одређени ниво дискриминације, иако је неетична, јер укључује контролу, доминацију и жртвовање појединца или целих друштвених група ипак води вештачку интелигенцију сазнањима о јединствености и непоновљивости људске личности, затим жртви, љубави и једнакости. Ова сазнања су у великој мери сагласна основним хришћанским учењима о оваплоћењу Сина Божјег у људском телу, те његовој јединственој и непоновљивој личности, затим о његовој крсној жртви и васкрсењу, као залогу васкрсења свих и уласка у живот вечни, и о једнакости људи која извире из чињенице да је човек створен по лику и подобију Божијем и да су сви људи подједнако позвани у заједницу са Богом. Ова сазнање даље утичу на изазове са којима ће се човечанство суочити када вештачка интелигенције достигне ниво суперинтелигенције. Тако уместо принципа контроле и доминације у односу између људи и супервештачке интелигенције требало да се промени принцип љубави и пожртвовања, исти онај којим се Бог Творац води у стварању творевине, а затим и у оваплоћењу унутар исте творевине као човек. Попут учења о Богу као Творцу о његовом оваплоћењу, и учење и конститутивног однос душе и тела у људском бићу може да отклони дилему о модалитетима и границима у спајању између људи и вештачке интелигенције. Слично томе, халкидонско диофизитско учење о две Христове природе у једној личности, може послужити као модел да се однос између људске природе и вештачке интелигенције не развија као однос двеју различитих природа, већ као однос слободних личности. Коначно, на питање усклађености циљева људског рода и вештачке интелигенције се може применити диотелитско учење о вољи, да би се показало да уколико вештачка интелигенција створена по принципу рационалности, онда и вештачка суперинтелигенција и човечанство одређују своје циљеви на основу исте рационалности и опредељења ка добру.

Литература

Agbada, Agbada (2020): Is the Killing of Qasem Soleimani by the United States of America Legal Under International Law? (January 7, 2020). Available at SSRN: <https://ssrn.com/abstract=3515524> or <http://dx.doi.org/10.2139/ssrn.3515524>

Agamben, Giorgio (2003): *Stato di eccezione*. Turin: Bollati Boringhieri Editore.

Appling, Scott Darren & Briscoe, Erica J. (2017): The Perception of Social Bots by Human and Machine. In: *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, Association for the Advancement of Artificial Intelligence (www.aaai.org), pp. 20-25.

Bathrellos, Demetrios (2012): *The Byzantine Christ*. Oxford University Press.

- Bostrom, Nick (2014): *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Cahour, Béatrice, and Forzy, Jean-François (2009): Does projection into use improve trust and exploration? An example with a cruise control system. *Safety Science* 47, 1260-1270.
- Chu, L. C., Anandkumar, A., Chang Shin, H., & Fishman, E. K. (2020): The Potential Dangers of Artificial Intelligence for Radiology and Radiologists. *Journal of the American College of Radiology*. doi:10.1016/j.jacr.2020.04.010
- Constas, Nicholas, (ed. and trans) (2014): *On the Difficulties in the Church Fathers: The Ambigua of Maximus the Confessor*. 2 vols. Cambridge, MA: Harvard University Press.
- Crenshaw, Kimberle (1989): Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, *The University of Chicago Legal Forum* 140, 139-167.
- Dalenberg, David J. (2018): Preventing Discrimination in the Automated Targeting of Job Advertisements, *Computer Law & Security Review* 34 (3), 615-627.
- DARPA (1983) *Strategic Computing: New Generation Computing Technology: A Strategic Plan for its Development and Application to Critical Problems in Defense*. Washington, DC: Defense Advanced Research Projects Agency (DARPA).
- Dillow, Clay (2016): All of These Countries Now Have Armed Drones, *Fortune*. 12 February, URL (consulted 19 November 2018): <http://fortune.com/2016/02/12/these-countries-have-armed-drones/>
- Dubhashi, Devdatt & Lappin, Shalom (2017): AI dangers, *Communications of the ACM*, 60 (2). 43-45.
- Elliott, Anthony (2018): Automated mobilities: From weaponized drones to killer bots. *Journal of Sociology*, 144078331881177. doi:10.1177/1440783318811777
- Ellis, Evelyn & Watson, Philippa (2012): *EU Anti Discrimination Law*. Oxford: Oxford University Press.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*. doi:10.1007/s11023-018-9482-5
- Geist, Edward Moore (2016): It's already too late to stop the AI arms race—We must manage it instead, *Bulletin of the Atomic Scientists*, 72(5), 318-321. doi:10.1080/00963402.2016.1216672
- González, Roberto J. (2017). Hacking the citizenry?: Personality profiling, “big data” and the election of Donald Trump. *Anthropology Today*, 33(3), 9–12. doi:10.1111/1467-8322.12348
- GPT-3, (8 September 2020). A robot wrote this entire article. Are you scared yet, human? *Guardian*. <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>
- Hoffman, R. R., Mueller, S. T., Klein, G. & Litman, J. (2018). Metrics for Explainable AI: Challenges and Prospects. *XAI Metrics*.
- Kendall, Frank (2014): Terms of Reference—Defense Science 2015 Summer Study on Autonomy. *Memorandum for Chairman, Defense Science Board*, www.acq.osd.mil/dsb/tors/TOR-2014-11-17-Summer_Study_2015_on_Autonomy.pdf.
- Kessler, Glenn (2019, March 22): The Iraq War and WMDs: An intelligence failure or White House spin? *The Washington Post*. <https://www.washingtonpost.com/politics/2019/03/22/iraq-war-wmds-an-intelligence-failure-or-white-house-spin/>
- Kurzweil, Ray 2005. *The Singularity Is Near* (Viking).
- Kurzweil, Ray 2012. *How to Create a Mind* (Viking Penguin).

- Levy, Antoine, 2015. "Γωνία: Looking into Corners of St Maximus' Cosmic Architecture." In: Antoine Lévy, Pauli Annala, Olli Hallamaa and Diana Kaley (eds.), *The Architecture of the Cosmos*. Helsinki: Luther- Agricola Society, 153–174
- Lipson, Hod and Melba Kurman (2016) *Driverless*. Cambridge, MA: MIT Press.
- Meisels, Tamar (2017). Targeted killing with drones? Old arguments, new technologies. *Philosophy & Society*, 29 (1), 1–152.
- Miller, Tim (2018): Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. doi: 10.1016/j.artint.2018.07.007
- Moravec, Hans (March 23, 2009). 'Rise of the Robots--The Future of Artificial Intelligence', *Scientific American*.
- Muir, Bonnie M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man–Machine Studies*, 27: 527–539.
- Muir, Bonnie M. (1994). Trust in automation Part 1: Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37, 1905–1922.
- Parfit, Derek. (1984). *Reasons and Persons*, Oxford: Clarendon Press.
- Remagnino, P., Shihab, A. L., & Jones, G. A. (2004). Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition*, 37(4), 675-689.
- Roussi, Antoaneta (2020): Resisting the rise of facial recognition, *Nature*, **587**: 350-353. doi:10.1038/d41586-020-03188-2
- Schmitt, Carl. (1985). *Political Theology: Four Chapters on the Concept of Sovereignty*, Cambridge, MA and London: The MIT Press.
- Styles, K. (2016) '40 Countries Are Working on Killer Robots and There's No Law to Say How We Use Them', The Next Web 22 January <https://thenextweb.com/us/2016/01/21/40-countries-are-working-on-killer-robots-and-theres-no-law-to-say-how-we-use-them>
- Taddeo, Mariarosaria (2018). The limits of deterrence theory in cyberspace. *Philosophy & Technology*, 31 (3), 339-355.
- Tegmark, Max (2017): *Life 3.0. Being Human in the Age of Artificial Intelligence*. Penguin.
- US Defense (n.d.) 'Unmanned Aircraft Systems (UAS)': <https://www.defense.gov/UAS/>
- Thomson, Judith Jarvis (2008): People and Their Bodies, in T. Sider, J. Hawthorne & D. W. Zimmerman (eds.), *Contemporary Debates in Metaphysics*, Blackwell, 155-176.
- Weber, Max (2004), *The Vocation Lectures*. Indianapolis/ Cambridge: Hackett Publishing Company 2004.
- Weyerer, Jan C. & Paul F. Langer, P. (2019). Garbage In, Garbage Out. *20th Annual International Conference on Digital Government Research on - Dg.o 2019*. doi:10.1145/3325112.3328220

**Ana Lipij
Ljubiša Bojić
Vladimir Cvetković**

**ARTIFICIAL INTELLIGENCE AND THE CHURCH FATHERS:
A PATRISTIC GUIDE TO THE CHALLENGES OF ARTIFICIAL INTELLIGENCE
AND SUPERINTELLIGENCE**

The article aims to explore the possibility of employing the doctrines developed by the Church Fathers in solving a number of contemporary questions in the application of the artificial intelligence. The article first points to employment of the AI in the surveillance and security systems, autonomous weapons, such as killer drones and robots, and public administrations. Next, it is argued that in spite of the ethically problematic AI employment, it is possible for AI to acquire some very positive knowledge about humankind, that is consistent with the principles of Christianity, and particularly patristic anthropology. From being employed in these three sectors, AI may acquire the knowledge about the principal dogmas of the Christian Church, namely the dogma of divine incarnation in the human form, the dogma of the death of God on the cross and his subsequent resurrection and the dogma about the equality of humankind in front of God. Finally, the article focuses on four artificial intelligence (AI) questions regarding: a) the consequences of rapid intelligence explosion or the fast take off, b) the ontological status of cyborgs and uploads, c) the future relations of artificial superintelligence and humanity, and d) the way of aligning the goals of super intelligent machines and humanity. In proposing solutions to these problems, the article applies patristic argumentation about God as Creator who creates humanity according to his image and likeness, about the soul-body relationship originating from the Origenist controversies, about the Chalcedonian definitions of two unconfused and undivided natures of Christ known as Dyophysitism, and about the Dyothelite doctrine of will as natural and personal capacity.