ANALIZA SLIK IN BESEDIL S PRISTOPI UMETNE INTELIGENCE

Priložnosti in dileme

Pregledni znanstveni članek | 1.02
Datum prejema: 27. 11. 2018

Izvleček: Napredek umetne inteligence, zlasti na področju nevronskih mrež, odpira mnogo etičnih dilem. Zmožnosti strojnega učenja umetnih nevronskih mrež članek prikaže na treh primerih. Prvi je diagnostika tumorjev z medicinskih slik, drugi določanje spolne usmerjenosti s fotografij obrazov in tretji analiza sentimenta na podlagi besedil. Izbrani primeri pokažejo prednosti novih tehnologij, opozorijo na njihove nevarnosti in osvetlijo možnosti uporabe za razumevanje velikih zbirk podatkov. Prispevek poudarja tudi etične dileme in vprašanja, ki se porajajo ob uporabi tehnik globokega učenja.

Ključne besede: umetna inteligenca, računalniški vid, obdelava naravnega jezika, analiza sentimenta, globoke nevronske mreže

Abstract: Recently, in the media, there has been a lot of news about artificial intelligence, its new capabilities in solving difficult problems, and related ethical dilemmas. The news mostly results from developments in the area of deep neural networks, which combine many layers of relatively simple computational units in an attempt to simulate the working of the biological brain. The machine learning capabilities of deep neural networks are demonstrated on three used cases: diagnostics of tumors from medical images, detection of sexual orientation from facial images, and sentiment analysis from texts. The chosen cases show the advantages of new technologies, their dangers, and the necessity of their use to understand large data sets. Furthermore, some ethical dilemmas and questions related to the use of deep learning are pointed out.

Keywords: artificial intelligence, computer vision, natural language processing, sentiment analysis, deep neural networks

Uvod

Nova odkritja in tehnologije so v mnogočem vplivale na človeka in družbo, zato so pogosto predmet poglobljenih antropoloških raziskav, za to področje pa se zanimajo tudi mediji. V zadnjih letih poročajo o razvoju umetne inteligence in opisujejo njene zmožnosti, hkrati pa opozarjajo na morebitne nevarnosti. Antropologi so v tem kontekstu primerni sogovorniki za področje etike pri uporabi novih tehnoloških rešitev, hkrati pa lahko pomembno prispevajo k razumevanju vpliva tehnologij na družbo ter snovanju novih rešitev, ki so smiselne za ljudi in koristne za varovanje okolja (Tisov idr. 2017).

Medijske reprezentacije in interpretacije umetne inteligence so večinoma senzacionalistične. Moderni preroki in (samo)promotorji, kot je poslovnež Elon Musk, napovedujejo skorajšnji zaton človeštva (Spletni vir 1). Z nastankom t. i. splošne umetne inteligence in samorazvijajočih se programov naj bi se pojavili novi in boljši sistemi, ki bodo razvijali še pametnejše stroje, ti pa bodo ljudi prej ko slej presegli. Ko bi dosegli to t. i. točko tehnološke singularnosti,[1] naj bi naš um in telo postala odvečna (glej Shanahan 2015). Ljudje bi se lahko rešili tako, da bi se združili s stroji in se naprej razvijali kot kiborgi. Večina tovrstnih domnev, ki temelji na razvoju splošne umetne inteligence, je prej rezultat skupinskih in individualnih strahov kot dejanske znanstvene, tehnološke in družbene realnosti. Nanje se je z znanstveno razlago odzval znani raziskovalec s področja robotike, Rodney Brooks (Spletni vir 2), ki je opisal nekaj napačnih načinov razmišljanja o umetni inteligenci in dejanske tehnološke omejitve pri njenem razvoju. Da bi se izognili pretiranemu senzacionalizmu, je treba razumeti temeljna načela umetne inteligence, poznati družbeni kontekst, v katerem je bila zasnovana, upoštevati pa je treba tudi kontekst, v katerem se bo uporabljala.

V članku prikazujeva tri primere uporabe strojnega učenja, obravnavava njihove predpostavke in rezultate ocenjujeva z antropološkega vidika. Izbrani primeri pokrivajo predvsem področje analize slik in besedil, kjer je bil v zadnjem času s pristopom globokih nevronskih mrež dosežen največji tehnološki napredek. S prvima primeroma računalniške analize slik oziroma fotografij želiva prikazati spekter možnih uporab tehnično enakih postopkov, razdelati etiko dela s podatki ter demonstrirati pomanjkljivost interpretacij, ki ne upoštevajo sodobnih družboslovnih ugotovitev. S tretjim primerom, s katerim predstavljava analizo sentimenta na veliki množici besedil, želiva pokazati možnosti

1 Tehnološka singularnost je hipoteza o umetni superinteligenci, ki naj bi nenadzorovano tehnološko rast s samoizboljšavami sprožila brez človeškega posredovanja.

* Ajda Pretnar, doktorska kandidatka na Univerzi v Ljubljani, Filozofska fakulteta, Oddelek za etnologijo in kulturno antropologijo, raziskovalka na Univerzi v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana; ajda.pretnar@fri.uni-lj.si.

** Marko Robnik-Šikonja, dr. računalništva in informatike, redni profesor, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana; marko.robnik@fri.uni-lj.si.

uporabe strojnega učenja tudi za antropološke namene. Trije izbrani primeri med sabo niso neposredno primerljivi, kar niti ni najin namen, saj želiva opozoriti predvsem na prednosti in pomanjkljivosti novih pristopov, utemeljenih na umetni inteligenci.

Izbira primerov

S področja računalniške analize slik sva izbrala dve raziskavi, ki za svoje delovanje uporabljata podobno tehnologijo strojnega učenja, vendar je tehnologija uporabljena za drugačen namen, analiza in interpretacija pa sta bistveno drugačni tudi z etičnega vidika. Razlog najine izbire je prav diametralno nasprotje družbenih implikacij njune uporabe. Prvi primer predstavlja uspešno uporabo umetne inteligence za prepoznavanje rakavih celic na histoloških slikah (Bejnordi 2017). Uspešnost prepoznavanja bolezenskih znakov na podlagi umetne inteligence v tem primeru presega zdravnike specialiste. Rutinska uporaba te tehnologije potencialno lahko reši številna življenja in pripomore h kakovostnejšemu in pravočasnemu zdravljenju. Umetna inteligenca, ki temelji na globokem učenju, pristopu, ki uporablja večnivojske nevronske mreže, se zato v medicinski diagnostiki in biomedicini vse bolj uspešno uporablja.

Za drugi primer sva izbrala razvrščanje fotografij obrazov glede na spolno usmerjenost fotografiranih oseb (Wang in Kosinski 2018). Gre za medijsko izpostavljeno in kontroverzno raziskavo, ki nas izzove, da premislimo o mejah uporabe tehnologije pa tudi o etiki, zasebnosti in digitalnih sledeh, ki jih namerno ali nenamerno puščamo bodisi pri lastni rabi digitalnih tehnologij (npr. s komentiranjem na družbenih omrežjih) bodisi pri širitvi uporabe tehnologij v družbi (npr. rutinsko snemanje z varnostnimi kamerami). Primer je zlasti zanimiv, ker predstavi sporno uporabo podobne tehnološke rešitve, kot je bila uporabljena v prvem primeru. V nadaljevanju kontrastno primerjava obe raziskavi in ju obravnavava tudi z etičnega vidika.

Tretji primer, ki ga obravnavava v članku, je analiza besedil. Pri tem predstaviva raziskavo na področju analize sentimenta, kjer iz zapisanega samodejno ugotavljava pozitivno, negativno ali nevtralno stališče pisca. Tovrstne raziskave omogočajo vpogled v veliko množico besedil, ki so jih ustvarili ljudje in jih objavili na spletu, njihova analiza pa bi bila brez tehnoloških pripomočkov skoraj nemogoča. Primer je lahko izhodišče za razmislek o novih možnostih, ki jih za antropološke raziskave strojno učenje ponuja zlasti na področjih digitalne, lingvistične ter zgodovinske antropologije.

Izbrani primeri prikažejo, kakšne so možnosti za uporabo globokega učenja. V temeljih so primeri podobni, imajo pa povsem različne družbene implikacije. Postopek, ki se uporablja za odkrivanje rakavih celic pri prognostiki raka na dojki, se lahko uporabi tudi za napovedovanje spolne usmerjenosti. Pri slednjem sta poleg tehnološke rešitve ključna še družbeni kontekst in razumevanje koncepta spolne usmerjenosti. Razdvojenost naravoslovja in tehniških ved ter družboslovja in humanistike ter pomanjkanje interdisciplinarnih raziskav simptomatično prispevata k zavajanju javnosti glede dejanskih zmožnosti umetne inteligence in njenih omejitev. V članku opozarjava na nevarnosti tovrstnega razdvajanja in prikaževa posledice vzajemnega nepoznavanja obeh področij.

Umetna inteligenca

Umetna inteligenca je zmožnost strojev (običajno računalnikov) za inteligentno delovanje. Natančneje to pomeni, da lahko pravilno prepoznajo vhodne podatke, se iz njih učijo in rezultate uporabijo za doseganje podanih ciljev. Če računalniku npr. podamo zdravstvene podatke o pacientih z rakom in pacientih brez raka, bo dobro naučeni model[2] lahko za novega pacienta na podlagi njegovih meritev napovedal, ali ima raka ali ne. Za področje umetne inteligence je pomembna tudi zamisel o splošni umetni inteligenci, torej sistemu, ki bi zmogel opravljati katerokoli nalogo tako inteligentno, kot jo opravlja človek. Zaenkrat smo še precej oddaljeni od kakršnekoli oblike splošne umetne inteligence, saj so uspešni modeli trenutno naučeni le za opravljanje specifičnih nalog (npr. napovedovanje raka, ne pa tudi ledvičnih kamnov).

Znanstveno področje umetne inteligence je v zadnjih letih doseglo velik napredek pri reševanju prej nerešljivih problemov. Največji napredek se je zgodil na področjih računalniškega vida (prepoznavanje obrazov je npr. že enakovredno človeškemu (Lu in Tang 2015)), igranja iger (leta 2015 je program premagal človeškega prvaka v igri go (Silver idr. 2017)) in razumevanja naravnega jezika (odlično strojno razpoznavanje govora in vse boljši rezultati strojnega prevajanja (Wu idr. 2016)). Ti uspehi so večinoma rezultat napredka na področju globokih nevronskih mrež (Goodfellow idr. 2016). Umetne nevronske mreže so modeli, sestavljeni iz velike zbirke povezanih računskih enot, imenovanih umetni nevroni, ki nekoliko posnemajo delovanje nevronov v možganih. Umetni nevroni podatke na svojem vhodu seštejejo in na izhod prenesejo nelinearno preslikavo vsote, kar celotni mreži omogoča, da simulira poljubno matematično preslikavo vhoda (Hornik 1991). V zadnjem času lahko raziskovalci v nevronske mreže učinkovito dodajajo vse več plasti »nevronov«. Tako imenovana globoka omrežja s številnimi sloji nevronov za uspešno učenje zahtevajo velike zbirke rešenih primerov in hitre ter vzporedno delujoče računalnike. Raziskovalci

2 Model je končni rezultat učenja, kjer učni algoritem iz vhodnih podatkov prepozna relevantne vzorce in na njihovi podlagi formulira matematično reprezentacijo problema. Model za prepoznavanje rakavih celic npr. najprej prejme množico slik, ki je označena z »rakavo« in »ne-rakavo«, nato pa se postopno nauči, kako obe vrsti celic čim bolj točno razlikovati. Tak model nato lahko za nove slike napove, ali je neka celica rakava ali ne.

imamo danes na razpolago oboje in se lahko zato lotimo mnogo težjih izzivov kot v preteklosti.

Metode umetne inteligence in podatkovnega rudarjenja

Umetna inteligenca je znanstveno področje, ki povezuje računalništvo, matematiko, psihologijo in nevroznanost. Njen cilj je ustvariti stroje, ki posnemajo človekove kognitivne funkcije, kot sta učenje ali reševanje problemov. Zgodovinsko gledano so uspehom umetne inteligence pogosto sledila razočaranja, ki so jih v veliki meri povzročile preveč optimistične napovedi tehnoloških vizionarjev. Tako je eden od ustanoviteljev področja umetne inteligence, Herbert Simon, v šestdesetih letih prejšnjega stoletja napovedal: »Stroji bodo v prihodnjih dvajsetih letih sposobni opraviti katerokoli delo, ki so ga zmožni ljudje« (Simon 1965). Marvin Minsky, pionir raziskav umetnih nevronskih mrež, je bil leta 1968 še bolj neposreden: »V eni generaciji bo problem snovanja umetne inteligence dejansko rešen« (po Crevier 1993). Oba sta se pozneje zavedala pretiranega optimizma in težavnosti splošne umetne inteligence.

Med pomembnimi novejšimi dosežki umetne inteligence je treba poudariti napredek na področju globokih nevronskih mrež, sistemov strojnega učenja, ki se na podlagi velike in običajno kompleksne množice podatkov sami naučijo reševati podano nalogo. V tehničnem smislu nevronsko mrežo sestavlja velika množica povezanih računskih enot, imenovanih umetni nevroni. Da bi tako omrežje naučili reševati probleme, mu moramo podati veliko že rešenih primerov. Mrežo denimo učimo z zbirko slik tkiva, v kateri so posamezne slike označene z diagnozo »rak« ali »ne-rak«. Vsako sliko iz zbirke pošljemo skozi mrežo, ki izračuna verjetnost, da je tkivo rakavo. Nato izračunamo razliko med napovedanim in pravilnim rezultatom ter uteži povezav med nevroni popravimo tako, da mreža vrača pravilnejše rezultate. Proces računanja napovedi in popravljanja povezav velikokrat ponovimo, vse dokler mreža pravilno ne napove večine slik. Tak postopek učenja imenujemo 'vzvratno razširjanje napake' (angl. *backpropagation*). Naučena umetna nevronska mreža bo sposobna opraviti delo, ki ga običajno opravljajo patologi, in bo na novih slikah zmožna prepoznati rakavo tkivo. Na podoben način se znanj in veščin priučimo tudi ljudje. Otroci se npr. igranja inštrumenta učijo z vajo in s ponavljanjem skladbe, dokler igranje ni dovolj dobro. Umetna nevronska mreža hrani naučene spretnosti, a shranjenega znanja, ki je dostopno v obliki omenjenih uteži na povezavah, ni enostavno razložiti in ga interpretirati v razumljivem jeziku.

Sodobne večnivojske nevronske mreže, ki so vse pomembnejše za področje umetne inteligence, so postale praktično uporabne šele po letu 2009, ko so raziskovalci z uporabo tisočev procesorjev na grafičnih karticah, ki se uporabljajo za njihovo učenje, dosegli dovolj hitro procesiranje. Naslednji pogoj za uspešnost globokega učenja so velike zbirke označenih podatkov. Raziskovalci so velike množice besedil in slik med drugim pridobili s pomočjo spleta, družbenih omrežij in Wikipedije. Tako so omogočili strojno prepoznavanje objektov na slikah, razpoznavanje govora in strojno prevajanje. Nekatere od teh računalniško izvedenih nalog sicer že dosegajo in presegajo človeka, a je kljub temu uspešno delovanje umetne inteligence še vedno omejeno le na posamezne naloge in nekatera področja. Za zdaj je npr. le malo napredka pri razumevanju pomena slik in besedil. Če preprosto karikaturo pokažemo naučeni nevronski mreži, bo prepoznala oblike, objekte in aktivnosti, ne bo pa razvozlala pomena, medsebojnih odnosov, zlasti pa ne humorja.

Nevronske mreže seveda niso edini model strojnega učenja. Obstaja vrsta različnih modelov, ki so pri reševanju nekaterih problemov uspešnejši, hitrejši oziroma potrebujejo manj učnih podatkov (Murphy 2012). Teorem zastonjskega kosila (Wolpert in Macready 1997), poenostavljeno povedano, celo trdi, da ne obstaja model strojnega učenja, ki bi bil primeren za reševanje vseh problemov. Številne raziskave so zato posvečene razvoju novih modelov in iskanju najustreznejših za posamezne pomembne probleme in aplikacije. Uspešne primere rabe umetne inteligence zasledimo npr. v medicini, farmaciji, financah, športu, telekomunikacijah, zavarovalništvu, bioinformatiki, marketingu, sociologiji, jezikoslovju itn.

Analiza slik z metodami globokega učenja

V pričujočem razdelku najprej predstavljava raziskave s področja medicine, v katerih je avtorjem s pomočjo globokega učenja uspelo doseči boljšo klasifikacijsko točnost histoloških slik, kot jo dosežejo zdravniki specialisti. Podobne tehnike globokega učenja za analizo slik nato predstavljava še na primeru raziskave, v kateri so s fotografij oseb, objavljenih na spletnem mestu za zmenke, skušali napovedati njihovo spolno usmerjenost. Ti raziskavi sva izbrala kot dva skrajna primera uporabe umetne inteligence. Po eni strani lahko s pomočjo modela za analizo slik žensk z rakom na prsih zdravniki uspešneje in hitreje prepoznajo rakava obolenja pri novih pacientkah, kar rešuje življenja in olajša medicinsko diagnostiko. Po drugi strani pa obstaja model, ki s fotografije obraza precej natančno prepozna spolno usmerjenost osebe, kar izzove kopico etičnih vprašanj tako o sami raziskavi kot tudi o možnostih uporabe (in zlorabe) tovrstnih modelov. Drugi primer podrobneje obravnavava z vidika metodologije in etike ter opozarjava na bistvene razlike med medicinskim in »fiziološkim« modelom.

Analiza medicinskih slik

Kot primer uspešne analize slik najprej nekoliko podrobneje predstavljava študijo uspešnosti diagnostike metastaz limfnih vozlišč pri ženskah z rakom dojk (Bejnordi idr. 2017). Avtorji so preizkusili uspešnost avtomatskih algo-

ritmov globokega učenja pri odkrivanju metastaz v tkivu limfnih vozlišč pacientk z rakom dojke in jo primerjali z uspešnostjo specialistov patologov. Študija temelji na znanstvenem izzivu Cancer Metastases in Lymph Nodes Challenge 2016 (CAMELYON16), ki je potekal novembra leta 2016 in pri katerem je sodelovalo 32 predlaganih algoritmov. Udeleženci izziva so iz univerzitetnih medicinskih centrov Redbound in Utrecht na Nizozemskem za učenje dobili 110 slik z metastazami in 160 slik brez njih. Zgrajene rešitve so bile ovrednotene na neodvisni testni množici 129 slik (49 z metastazami in 80 brez). Isto testno množico je ocenilo tudi 11 nizozemskih patologov, ki so bili v simulaciji običajnega delovnega procesa časovno omejeni na dve uri, in en patolog brez časovnih omejitev, ki je za natančno analizo porabil 30 ur. Zanesljivost svojih napovedi so patologi ocenili z ocenami zanesljivo normalno, verjetno normalno, nedoločno, verjetno tumor in zanesljivo tumor. Študija primerja uspešnost algoritmov in patologov pri dveh nalogah, identifikaciji posameznih metastaz in binarni klasifikacijski nalogi (prisotnost ali odsotnost metastaz na celotni sliki). Točnost klasifikacije študija poroča z mero AUC (Area Under the ROC curve), kjer vrednost 0,81 na primer pomeni, da je pri dveh naključno izbranih slikah z in brez metastaz klasifikator pravilno izbral sliko z metastazami v 81 odstotkih primerov.

Najboljši algoritem je uspešno identificiral 80,7 odstotka metastaz (nobene napačno), patolog brez časovnih omejitev pa 72,4 odstotka (prav tako nobene napačno). Pri binarni nalogi je bila klasifikacijska uspešnost najboljšega algoritma (AUC = 0,994) bistveno višja od povprečja časovno omejenih patologov (AUC = 0,810), ki so sicer metastaze prepoznavali z uspešnostjo med 0,738 in 0,884. Uspešnost najboljšega algoritma je povsem primerljiva s patologom brez časovne omejitve (AUC = 0,966). Najboljših sedem algoritmov je statistično značilno preseglo klasifikacijsko točnost časovno omejenih patologov. Avtomatski pristopi so uporabljali iz literature znane arhitekture globokih nevronskih mrež (GoogLeNet, ResNet-101, VGG-16).

V primerjavi z dejanskim stanjem ima študija nekatere omejitve. Za vsakega pacienta je bila npr. na voljo le ena slika, medtem ko lahko v bolnišnični praksi patologi v dvomljivih primerih zahtevajo dodatne slike; podobno tudi časovna omejitev na dve uri za 129 slik, čeprav je dokaj realna, ne odraža dejanskega delovnega postopka v bolnišnici. Kljub temu pa ima tehnologija globokih nevronskih mrež očitno velik potencial za praktično uporabo v medicinski diagnostiki. Sorodne študije kažejo na njihovo uspešnost pri odkrivanju diabetične retinopatije s slik očesnega ozadja (Gulshan idr. 2016), kožnega in drugih vrst raka (Esteva idr. 2017), v klinični praksi pa so modeli strojnega učenja vse bolj koristen pripomoček za uspešno in hitro diagnostiko nekaterih bolezni.

Etične implikacije te študije so povezane zlasti z zbiranjem občutljivih zdravstvenih podatkov. Podatki so bili pridobljeni na javnem natečaju, pri čemer so organizatorji navedli tudi njihov vir (dva nizozemska univerzitetna medicinska centra). Univerzitetna medicinska centra sta pridobila soglasje pacientk, slike pa so bile posredovane brez informacije o pacientkah, le s podatkom o njihovem zdravstvenem stanju (z metastazami ali brez). Cilj študije je bil omogočiti boljšo in hitrejšo diagnostiko in po najini oceni ustreza tudi antropološki raziskovalni etiki, ki upošteva človekovo dostojanstvo, avtonomijo, zaščito, maksimizira koristi in minimizira škodo ter spodbuja spoštovanje in pravičnost (Spletni vir 3; Halford 2017).

Napovedovanje spolne usmerjenosti

Wang in Kosinski (2018) sta zmožnosti globokega učenja demonstrirala na kontroverznem problemu določanja spolne usmerjenosti s fotografij obrazov. Študija odpira številna metodološka in etična vprašanja; del teh vprašanj obravnavava v nadaljevanju. Avtorja sta poskušala na nekatere ugovore in najbolj pereča vprašanja odgovoriti v posebnem spletnem dokumentu (Spletni vir 4). Med drugim poudarjata, da številne vlade in podjetja globoke nevronske mreže že uporabljajo za analizo obrazov in se dobro zavedajo zmožnosti te tehnologije za identifikacijo oseb in njihovih občutljivih lastnosti. Ker pa se tega ne zavedajo splošna javnost, zakonodajalci in znanstveniki, jih avtorja s svojo raziskavo želita ozavestiti o realnosti tveganj in jih pojasniti; istospolna usmerjenost je mnogokje po svetu še vedno preganjana in celo kazniva.

Omenjena raziskava preverja hipotezo o informativnosti fiziognomije, ki trdi, da obstajajo povezave med našim značajem in videzom. Hipoteza se pri tem najbolj osredotoča na obliko in poteze obraza, pa tudi na roke in druge dele telesa. Začetki te hipoteze segajo v antiko, ko sta se fiziognomiji posvečala Aristotel in Pitagora. Hipoteza je bila zavrnjena in tabuizirana v 20. stoletju (Livingstone 1962; Molnar 1983; Smedley 2012), vendar obstajajo nekatere raziskave, ki kažejo, da ljudje z majhno, vendar značilno višjo verjetnostjo od naključne z obraza prepoznavamo nekatere lastnosti, kot so spol, starost, čustvena stanja, osebnostne lastnosti, spolna usmerjenost, politično prepričanje ali celo možnost zmage na volitvah. Znanstveni temelj te hipoteze je, da na obrazne poteze, npr. lahko vplivajo predporodna in poporodna izpostavljenost hormonom, okoljski in genetski vplivi itd. Temelj za prepoznavanje spolne usmerjenosti s fotografij obrazov je prav t. i. predporodna hormonska teorija, ki trdi, da istospolno orientiranost povzroča premajhna izpostavljenost moškega ali prevelika izpostavljenost ženskega fetusa moškim spolnim hormonom; isti hormoni naj bi povzročili tudi spolno diferenciacijo obraza, s čimer bi se pokazala spolna usmerjenost osebe. Po tej teoriji bi imeli istospolni moški manjše čeljusti in lica, tanjše obrvi, daljši nos in večje čelo (obratno za ženske); spolno atipični naj bi bili tudi oblačenje, frizura, nega obraza itd.

Hipotezo je v praksi skušala preveriti omenjena študija. Vir informacij za določanje spolne usmerjenosti z obraznih slik so bile fotografije s spletnega mesta za zmenke. Avtorja sta spolno usmerjenost oseb določila na podlagi zabeleženega spola in spola iskanega partnerja oziroma partnerice. V začetni zbirki sta zbrala 130.741 fotografij 36.630 moških in 170.360 slik 38.593 žensk, starih med 18 in 40 let. V zbirki je bilo enako število homo- in heteroseksualnih oseb. Po nadaljnjem čiščenju podatkov s programom Face++ in s pomočjo množičenja na portalu Amazon Mechanical Turk (AMT), pri čemer sta izločila vse premajhne, nagnjene in nestandardne fotografije ter fotografije vseh nebelopoltih oseb, jima je ostalo 35.326 fotografij s približno enako zastopanimi moškimi in ženskami ter enako zastopanimi isto in različno spolno usmerjenimi osebami. Fotografije sta skalirala na velikost 224 x 224 točk, za vsako osebo pa sta v zbirki obdržala od ene do pet fotografij.

Globoko nevronsko mrežo sta avtorja uporabljala za ekstrakcijo značilk z obraznih slik, in sicer sta uporabila že vnaprej naučeno nevronsko mrežo VGG-Face, ki je namenjena problemu prepoznavanja obrazov in je naučena na 2,6 milijona slik. Mreža je z vsake fotografije generirala 4096 značilk. Po redukciji značilk s postopkom SVD sta ohranila 500 značilk. Na njih sta kot klasifikator uporabila logistično regresijo z regularizacijo LASSO. Rezultate sta statistično ovrednotila z 20-kratnim prečnim preverjanjem. Kot klasifikacijsko točnost poročata mero AUC (Area Under the ROC curve), kjer vrednost 0,81 pomeni, da je med naključno izbranimi pari homo- in heteroseksualnih oseb klasifikator pravilno ugotovil spolno usmerjenost v 81 odstotkih primerov (interpretacija je enaka kot pri Wilcoxonovem testu s predznačenimi rangi). V raziskavi sta avtorja skušala empirično odgovoriti na naslednja vprašanja:

1. Je mogoče na podlagi obraznih slik med seboj ločiti homoseksualne moške od heteroseksualnih in homoseksualne ženske od heteroseksualnih?
2. Kateri deli obraza so pomembni za ločevanje?
3. Kakšen je tipičen obraz isto- in različno spolno usmerjenih oseb?
4. Kako dobro spolno usmerjenost prepoznavamo ljudje?

Rezultati so pokazali, da pri eni fotografiji obraza na osebo nevronska mreža za fotografije moških doseže AUC = 0,81, za fotografije žensk pa AUC = 0,71. Pri petih fotografijah vrednosti pri moških narastejo na 0,91 in pri ženskah na 0,83. Kot pomembni deli obraza za ločevanje so se s pomočjo skrivanja posameznih obraznih delov in preverjanja uspešnosti klasifikacije izkazali usta, nos in brada. Za merjenje uspešnosti prepoznavanja ljudi sta avtorja v raziskavi preizkusila 35 človeških ocenjevalcev, ki so za moške dosegli AUC = 0,61, za ženske pa 0,54. Avtorja menita, da rezultati podpirajo veljavnost predporodne hormonske teorije, po kateri imajo homoseksualni moški in ženske za svoj spol netipično obrazno morfologijo in izraz, stil frizure itd.

Rezultati te raziskave imajo lahko pomembne posledice za razumevanje spolne usmerjenosti in omejenosti človeške zmožnosti prepoznavanja obrazov. Ob njej pa se postavljajo nekatere metodološke omejitve. Prvič, fotografije niso nujno reprezentativne, saj jih kurirajo uporabniki. Zbirka vsebuje samo podatke odprto istospolno usmerjenih ljudi, ki so svoje fotografije objavili na spletnem mestu za zmenke. Zelo verjetno je, da odprto istospolno usmerjeni zunanji izgled prilagodijo lažji identifikaciji pri iskanju partnerjev. Fotografije splošne populacije bi bile lahko drugačne. Drugič, na fotografijah so bili samo belopolti kandidati. Ta omejitev je stvar vnaprejšnje odločitve avtorjev študije, da bi zmanjšali potreben napor pri zbiranju in označevanju slik. Študije prepoznavanja obrazov sicer kažejo, da je uspešnost prepoznavanja ob zadostni učni množici enaka za vse barve polti, vendar pa ni podatkov o uspešnosti prepoznavanja spolne usmerjenosti na celotnem spektru morfoloških tipov. Tretjič, slabša uspešnost ljudi, ki naj bi demonstrirala superiornost tehnik umetne inteligence v primerjavi z ljudmi pri razpoznavanju informacij iz obrazov, je lahko posledica tega, da ljudje običajno nimamo na razpolago tako velike in na tak način označene učne množice ter da se za spolno usmerjenost neznancev v vsakdanjem življenju ne zanimamo. Četrtič, spolna identiteta je v tej študiji razumljena binarno, izpuščeni pa so primeri kompleksnejših identifikacij. Raznolikost spolnih identitet je, skratka, povsem poenostavljena.

Opisana študija tako ne sporoča nujno, da lahko algoritmi uspešno prepoznajo spolno usmerjenost ljudi na podlagi njihovih profilnih fotografij, temveč predvsem, da je mogoče z globokim učenjem doseči zadovoljivo uspešnost prepoznavanja spolne identifikacije na zelo omejenem vzorcu in z nekaj pomembnimi predpostavkami. Kljub temu da je primer zanimiv opomnik na potencialne zlorabe tehnologije, izpušča ključne sociološke ugotovitve s področja spolnih študij, ki bi raziskavi dale pravilen kontekst in opozorile na omejitve aplikacije modela. Z etičnega vidika ta študija posega tudi v dostojanstvo ljudi in ne poskrbi za minimiziranje škode, napačno razumljeni rezultati imajo namreč lahko resne posledice za skupnost LGBTIQ, hkrati pa tudi nasploh za sodobno družbo, ki naj bi bila utemeljena na pravičnosti in enakopravnosti.

Analiza besedil s strojnim učenjem

Čeprav za velike zbirke označenih besedil v zadnjem času prevladuje pristop z globokimi nevronskimi mrežami, so pri manjših količinah podatkov mnogokrat uspešnejše klasične metode strojnega učenja, kot so metoda podpornih vektorjev, naivni Bayesov klasifikator ali logistična regresija. V nadaljevanju predstavljava problem analize določanja sentimenta iz besedil. Najin namen je predstaviti eno

pogostih aplikacij strojnega učenja na besedilnih podatkih, ki je lahko zanimiv metodološki pristop za družboslovne in humanistične raziskave.

Analiza sentimenta

Analiza 'mnenj v besedilih' (angl. *opinion mining*), v ožjem kontekstu poimenovana tudi 'analiza sentimenta' (angl. *sentiment analysis*), je eno od področij tekstovnega rudarjenja (Liu in Zhang 2012). Ukvarja se z odkrivanjem piščevega mnenja o predmetu pisanja. Večinoma se uporablja polarna analiza, kar pomeni, da določamo pozitivno ali negativno mnenje. Naloga za avtomatske sisteme ni enostavna, saj je treba iz besedila izluščiti bistveno semantično informacijo, pri čemer so težave (večkratno) zanikanje, sarkazem, dvoumnost besedila, kontekstna odvisnost rabe besed itd. Pravilno določanje sentimenta je koristno za številne namene, npr. za napovedovanje uspešnosti izdelkov, izidov volitev ali v sociooloških raziskavah. Zaradi vse širšega izražanja mnenj na internetu preko spletnih forumov, komentarjev novic, tvitov ter recenzij izdelkov in storitev postaja avtomatska analiza mnenj pomemben raziskovalni pripomoček.

Analizo sentimenta lahko izvajamo na različnih ravneh, od najbolj splošne, tj. na ravni celotnega besedila, do posameznih vidikov oziroma značilnosti, ki se v besedilu pojavijo (npr. piščevo mnenje o porabi goriva testnega avtomobila). Pri slednji ravni je velika težava že identifikacija entitet in vidikov. V nadaljevanju prikazujeva primer analize na ravni celotnega besedila. Takšni analizi sentimenta pravimo tudi klasifikacija sentimenta, saj podobno kot pri drugih nalogah klasifikacije besedil (npr. klasifikacija dnevnih novic v skupine, kot so gospodarstvo, šport ipd.) besedilo klasificiramo v eno od danih kategorij.

Pri analizi sentimenta sta se najbolj uveljavila dva pristopa; prvi je leksikalni, drugi pa temelji na strojnem učenju. Pri leksikalni metodi potrebujemo enega ali več leksikonov sentimenta, ki vključujejo besede in fraze s pozitivno in z negativno konotacijo. Pristop prešteje te besede in fraze v besedilu, ki ga želimo klasificirati. Če prevladujejo besede z negativno konotacijo, besedilo označi kot negativno, sicer kot pozitivno. Težava pristopa je, da je treba pripraviti leksikon, poleg tega pa se izrazoslovje postopno spreminja. Nekatere besede imajo različen sentiment v različnih kontekstih (npr. beseda majhen se pri opisu zaslona mobilnega telefona šteje kot negativna oznaka, pri opisu pomnilniškega ključka pa kot pozitivna), zato je ta pristop v praksi v zadnjem času skoraj vedno združen s strojnim učenjem. Pri pristopu s strojnim učenjem s pomočjo učne množice besedil, ki jim priredimo eno od sentimentnih kategorij (npr. pozitivno ali negativno), tvorimo učno množico, na podlagi katere zgradimo klasifikacijski model. Kljub temu da sentimentni leksikon za analizo s strojnim učenjem ni nujen, so raziskave pokazale, da je dober leksikon koristen in da izboljša klasifikacijsko točnost (Hu in Liu 2004). Za slovenščino je splošni sentimentni leksikon na podlagi leksikona (Hu in Liu 2004) večinoma ročno sestavil Kadunc (2016). Leksikon obsega 2646 pozitivnih in 6689 negativnih besed.

Kadunc in Robnik-Šikonja (2016) sta primerjala uspešnost napovedovanja sentimenta spletnih uporabniških vsebin. Sestavila sta korpus slovenskih spletnih komentarjev, zbranih s spletnih strani štirih medijev (RtvSlo, 24 ur, Finance in Reporter) in klasificiranih v štiri tematska področja (gospodarstvo, politika, šport in drugo). Korpus sestavlja 4777 spletnih komentarjev, ki so ročno označeni s trivalentnim sentimentom (pozitivno, negativno in nevtralno). Ujemanje treh označevalcev, merjeno z mero Fleiss Kappa, je 0,38. Ta mera ocenjuje, koliko se ujemanje označevalcev razlikuje od ujemanja, ki bi ga dosegli z naključnim ocenjevanjem. Vrednost 1 pomeni, da se odločevalci popolnoma ujemajo, vrednost 0 pa, da ni razlike z naključnim dodeljevanjem ocen. Kot pri drugih tovrstnih študijah je celoten postopek večfazen in sestavljen iz priprave podatkov, tokenizacije (razbitja na besede oz. druge osnovne enote, npr. emotikone), ekstrakcije značilk, testiranja uspešnosti različnih klasifikatorjev in statistične analize rezultatov z uporabo npr. prečnega preverjanja.

Kot najuspešnejši pristop za napovedovanje sentimenta se je izkazalo strojno učenje v kombinaciji s sentimentnim leksikonom. Ta je upoštevan tako, da se med značilke vključi tudi število besed, ki nastopajo v sentimentnem leksikonu (posebej število pozitivnih in negativnih besed). S tem pristopom je bilo sentiment mogoče klasificirati s približno 65-odstotno klasifikacijsko točnostjo, kar je glede na tri razrede in dokaj šibko ujemanje treh človeških označevalcev dober rezultat in primerljiv z rezultati analize sentimenta v drugih jezikih.

Naučene klasifikatorje sentimenta je mogoče uporabiti na velikih zbirkah besedil, ki so za ročno analizo preobsežne. Tako dobimo orodje, ki lahko poda odgovore na mnoga zanimiva raziskovalna in komercialna vprašanja. Še več, čeprav nimamo na razpolago označenih zbirk sentimenta, da bi modele lahko naučili strojnega učenja, nam obstoj sentimentnih leksikonov zagotavlja solidne klasifikacijske točnosti. To velja tudi za analizo sentimenta v slovenščini. Antropologi lahko tovrstne modele uporabljajo za hitro in učinkovito analizo družbenih omrežij, spletnih forumov, arhivskega gradiva ter navsezadnje tudi transkribiranih intervjujev. Poleg analize sentimenta so na podobnih principih utemeljene tehnike prepoznavanja tem v besedilih, gručenja dokumentov po podobnosti, luščenja informacij in podobno. Celotni spekter metod strojne analize besedil je lahko močno orodje raziskovalcev ne zgolj v jezikoslovju, temveč v celotni humanistiki.

Problem interpretacije rezultatov

»Ena mojih znanstvenih dolžnosti je, da če poznam nekaj, kar bi lahko obvarovalo ljudi pred škodo, moram to objaviti,« je izjavil Michal Kosinski, soavtor sporne študije o spolni usmerjenosti (po Spletni vir 5). Namen Kosinskega in Wanga pri objavi članka je bil, po njunih besedah, na preprostem primeru predstaviti potencialne nevarnosti uporabe umetne inteligence. Z uporabo javno dostopnih orodij, kot je Face++, in z uveljavljeno metodologijo strojnega učenja sta pokazala, da lahko model z 81-odstotno natančnostjo loči med samoopredeljenim svetlopoltim homoseksualnim ameriškim moškim, starim med 18 in 40 let, ki išče istospolnega partnerja na spletnih portalih, ter statistično podobnim heteroseksualnim moškim.[3] Ta poved je namenoma zapisana kompleksno, saj so kompleksni tudi zaključki njune raziskave.

Medtem ko ni nujno oporekati jasno definirani metodologiji, je treba temeljito premisliti in komentirati kontekst te raziskave. Nekaj omejitev je že navedenih v zgornjem poglavju pričujočega prispevka, obstaja pa še kopica dodatnih opozoril. Starostni vzorec preučevanih oseb je izrazito ozek in ne zajema celotne populacije. Spolna usmerjenost je definirana binarno, brez možnosti kompleksnih identifikacij (Middleton 2002). Uporabljeni model, kljub poskusom razlage vsebine, ni razumljiv in ne omogoča interpretiranja in preverjanja rezultatov. Študije 'pridobitvenih funkcij' (angl. *gain-of-function*) so sicer zanimive, saj simulirajo ekstremne situacije, ki jih, preden dobijo svoje mesto v dejanskem življenju, lahko premislimo. To naj bi bil tudi motiv za omenjeno raziskavo. Vendar pa ima vsak eksperiment, ki naslavlja družbene odnose in razmerja moči, tudi svoj družbeni kontekst. Avtorja uspešno uporabita standardne postopke iz raziskav o strojnem učenju, zanemarita pa pomemben del družboslovnih in humanističnih znanj, ki obravnavajo spolne identifikacije, morfološko pestrost človeške vrste ter samopredstavljanje na spletu. Brez tovrstnih znanj so študije kontekstualno okrnjene in v družbi dejansko niso uporabne (Dourish in Gómez Cruz 2018).

Da so modeli globokega učenja močno orodje, ki ga lahko uporabljamo na različne načine, sva pokazala na treh primerih. Kosinski in Wang sta pokazala, da je, kot pri vseh orodjih, zloraba možna tudi pri teh modelih. Po drugi strani pa lahko enaki analitični pristopi rešujejo življenja (Bejnordi idr. 2017) ter pomagajo pri analizi in razumevanju velikih množic podatkov (Kadunc in Robnik-Šikonja 2016). Globoki nevronski modeli se navsezadnje lahko uporabljajo tudi za klasifikacijo kulturne dediščine (Pretnar idr. 2018). Spekter uporabe teh modelov je širok in le podrobno poznavanje družbenega konteksta in kulturnih specifik lahko posamezni model umesti v znanstveno razpravo ter ga prenese v prakso.

3 Z manjšo točnostjo podobno velja tudi za ženske.

Sklep

V prispevku predstavljava sodobne pristope s področja umetne inteligence, ki temeljijo na globokih nevronskih mrežah, in jih podrobneje opisujeva na treh primerih. S prvim primerom predstavljava uporabo avtomatskega prepoznavanja metastaz z medicinskih slik, kjer so metode globokih nevronskih mrež uspešnejše od zdravnikov patologov. Tovrstna raba lahko pomembno pripomore k uspešnosti zdravljenja in kakovosti življenja bolnikov. Drugi primer je določanje spolne usmerjenosti s fotografij obrazov, ki kaže, da lahko tehnologija resno krši pravice do zasebnosti in zlorablja občutljive informacije. Tretji primer, uporaba tehnik strojnega učenja na besedilih za analizo sentimenta spletnih uporabniških vsebin, pa prikazuje, kako je v času, ko ustvarjamo vse več velikih podatkovnih zbirk, uporaba sodobnih tehnologij nujna za uspešno raziskovalno delo in razumevanje načinov življenja.

Modeli globokega učenja in tudi drugi modeli strojnega učenja so pri obdelavi »naravoslovnih« podatkov, npr. medicinskih, bioloških ali genetskih, redko problematični. Običajno so neproblematični in zanimivi tudi v finančnih domenah, fiziki (Huertas-Company idr. 2018), geologiji, gradbeništvu, strojništvu, glasbi (Rehmeyer 2007) itd. Ko pa so modeli strojnega učenja naučeni na »družboslovnih« podatkih in imajo za cilj razumevanje in modeliranje kompleksne družbene realnosti, njihova analiza zahteva temeljit premislek o izvoru podatkov in njihovem širšem družbenem kontekstu. Ti podatki pogosto vsebujejo pomembne predpostavke o naravi družbenega, ki jih je pred zanesljivo uporabo različnih modelov treba najprej razkriti in upoštevati (Wagner-Pacifici, Mohr in Breiger 2015).

Za kakovostno interpretacijo rezultatov bo treba, zlasti pri modeliranju družbenih fenomenov, izsledke strojnega učenja obogatiti s sociološkimi in z antropološkimi znanji. Pogoj za to so tehnološko vešči družboslovci, ki razumejo delovanje algoritmov, ali tesno sodelovanje ekspertov s področja umetne inteligence in družboslovja. Samoumevno se zdi, npr., da bodo rezultate uporabe umetne inteligence na področju medicine interpretirali zdravniki. Zakaj bi bilo na področju proučevanja družbe drugače? S celostnim in z interdisciplinarnim pristopom se lahko izognemo senzacionalizaciji umetne inteligence ter naučene modele interpretiramo na način, ki upošteva kontekst podatkov in možnosti praktične uporabe modelov, hkrati pa opozarja na etično komponento vsake raziskave, ki obravnava družbene fenomene.

Literatura

BEJNORDI, Babak idr.: Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 318 (22), 2017, 2199–2210.

CREVIER, Daniel: *AI: The Tumultuous Search for Artificial Intelligence*. New York, NY: BasicBooks, 1993.

DOURISH, Paul in Edgar Gomez Cruz: Datafication and Data Fiction: Narrating Data and Narrating with Data. *Big Data & Society*, 2018.

ESTEVA, Andre idr.: Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature* 542, 2017, 115–118.

GOODFELLOW, Ian, Yoshua Bengio in Aaron Courville: *Deep Learning.* Cambridge, MA: MIT press, 2016.

GULSHAN, Varun idr.: Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 316 (22), 2016, 2402–2410.

HALFORD, Susan: The Ethical Disruptions of Social Media Data: Tales from the Field. V: Kandy Woodfield (ur.), Advances in Research Ethics and Integrity (Volume 2): *The Ethics of Online Research*. United Kingdom: Emerald Publishing Limited, 2017, 13–25.

HORNIK, Kurt: Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks* 4 (2), 1991, 251–257.

HU, Minqing in Bing Liu: Mining Opinion Features in Customer Reviews. V: Anthony G. Cohn (ur.), *Proceedings of the 19th National Conference on Artificial intelligence (AAAI'04)*. AAAI Press, 2004, 755–760.

HUERTAS-COMPANY, Marc idr.: Deep Learning Identifies High-z Galaxies in a Central Blue Nugget Phase in a Characteristic Mass Range. *Astrophysical Journal* 858 (2), 2018, 114.

KADUNC, Klemen: *Določanje sentimenta slovenskim spletnim komentarjem s pomočjo strojnega učenja*. Diplomsko delo. Ljubljana: Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, 2016.

KADUNC, Klemen in Marko Robnik-Šikonja: Analiza mnenj s pomočjo strojnega učenja in slovenskega leksikona sentimenta. V: Tomaž Erjavec in Darja Fišer (ur.), *Konferenca Jezikovne tehnologije in digitalna humanistika*. Ljubljana: Znanstvena založba Filozofske fakultete, 2016, 83–89.

LIU, Bing in Lei Zhang: A Survey of Opinion Mining and Sentiment analysis. V: Charu C. Aggarwal in ChangXiang Zhai (ur.), *Mining Text Data*. Boston, MA: Springer, 2012, 415–463.

LIVINGSTONE, Frank B.: On the Non-Existence of Human Races. *Current Anthropology* 3 (3), 1962, 279–281.

LU, Chaochao in Xiaoou Tang: Surpassing Human-Level Face Verification Performance on LFW with GaussianFace. V: Blai Bonet in Sven Koenig (ur.), *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. Palo Alto, California: AAAI Press, 2015, 3811–3819.

MIDDLETON, DeWight R.: *Exotics and Erotics: Human Cultural and Sexual Diversity*. Long Grove, IL, ZDA: Waveland PressInc, 2002.

MOLNAR, Stephen: *Human Variation: Races, Types, and Ethnic Groups*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

MURPHY, Kevin P.: *Machine Learning: A Probabilistic Perspective.* Cambridge, MA, ZDA: MIT Press, 2012.

PRETNAR, Ajda idr.: Power of Algorithms for Cultural Heritage Classification: The Case of Slovenian Hayracks. V: Daria Spampinato (ur.), *AIUCD 2018: Patrimoni culturali nell'era digitale: Memorie, culture umanistiche e technologie*. Bologna: Università di Bari Aldo Moro, 2018, 212–215.

REHMEYER, Julie J.: The Machine's Got Rhythm: Computers are Learning to Understand Music and Join the Band. *Science News* 171 (16), 2007, 248–250.

SHANAHAN, Murray: *The Technological Singularity.* Cambridge, MA, ZDA: MIT Press, 2015.

SILVER, David idr.: Mastering the Game of Go without Human Knowledge. *Nature* 550, 2017, 354–359.

SIMON, Herbert A.: *The Shape of Automation for Men and Management*. New York: Harper and Row, 1965.

SMEDLEY, Audrey in Brian D. Smedley: *Race in North America: Origin and Evolution of a Worldview*. New York: Westview Press, 2012.

TISOV, Ana idr.: People-Centred Approach for ICT Tools Supporting Energy Efficient and Healthy Behaviour in Buildings. V: Zia Lennard (ur.), *Proceedings of the Sustainable Places 2017 (SP2017) Conference*. Middlesbrough, UK: MDPI, 2017, 675.

WAGNER-PACIFICI, Robin, John W. Mohr in Ronald L. Breiger: Ontologies, Methodologies, and New Uses of Big Data in the Social and Cultural Sciences. *Big Data & Society*, 2015, 1–11.

WANG, Yilun in Michal Kosinski: Deep Neural Networks are More Accurate than Humans at Detecting Sexual Orientation from Facial Images. *Journal of Personality and Social Psychology* 114 (2), 2018, 246–257.

WU, Yonghui idr.: Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv:1609.08144v2*.

WOLPERT, David H. in William G. Macready: No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation* 1 (1), 1997, 67–82.

Spletni viri

Spletni vir 1: CLIFFORD, Catherine: Elon Musk: 'Mark my words – A.I. is Far More Dangerous than Nukes, 13. 3. 2018; https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html, 24. 9. 2018.

Spletni vir 2: BROOKS, Rodney: The Seven Deadly Sins of Predicting the Future of AI, 2017; https://rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/, 19. 9. 2018.

Spletni vir 3: MARKHAM, Annette, Elizabeth Buchanan in AoIR Ethics Working Committee: Ethical Decision-making and Internet Research: Version 2.0, 2012; http://www.aoir.org/reports/ethics2.pdf, 12. 2. 2019.

Spletni vir 4: KOSINSKI, Michal: A Word from the Coauthor about a Recent Peer-Reviewed Study, 13.09.2017; https://www.gsb.stanford.edu/newsroom/school-news/word-coauthor-about-recent-peer-reviewed-study, 20. 9. 2018.

Spletni vir 5: LEVIN, Sam: LGBT Groups Denounce 'Dangerous' AI that Uses Your Face to Guess Sexuality, 9. 9. 2017; https://www.theguardian.com/world/2017/sep/08/ai-gay-gaydar-algorithm-facial-recognition-criticism-stanford, 24. 9. 2018.

Analysis of Images and Text with AI Methods:
Opportunities and Dilemmas

In recent years, there has been a lot of mentioning in the media of artificial intelligence, both about its capabilities and potential dangers. The greatest progress was made in computer vision, playing games and understanding the natural language. These successes are mostly the result of progress in deep neural networks, which are large collections of connected computational units called artificial neurons that model neurons from the brain.
The first example of a successful use of artificial intelligence presented in the paper is identification of cancer cells in histopathology images. The accuracy of diagnostic assessment in this case surpasses medical specialists, meaning that a routine use of such technology potentially saves lives. Such models are a useful diagnostic tool in clinical praxis.
A contrasting example is detection of sexual orientation from photographs. The research tests the hypothesis of physiognomy, which claims there is a connection between character and appearance, where the focus is on the shape and features of the face. As a source of data the authors collected images from a dating website. They used a deep neural network to extract features from the images and trained a simple model to discriminate between hetero- and homosexual individuals. They achieved a high accuracy, but it is important to note that every experiment dealing with social relations has its particular context. The authors used standard computer science methodology, but neglected the findings from social sciences and the humanities.
Finally, the paper presents research on sentiment classification as an example of text analysis. Such research allows an efficient analysis of large data sets. Anthropologists can use these models for analysing social media, web fora, archival materials and interview transcripts.
The above examples are used to showcase different ways deep models can be used. Just like any tool, deep models can also be abused and misinterpreted. For a quality interpretation of the results, the authors argue for the enrichment of findings from machine learning through socio-anthropological knowledge, especially when modelling social phenomena.