

Veliki podaci u psihologiji ličnosti

Zrinka Bakić

Filozofski fakultet u Zagrebu
Odsjek za psihologiju
ORCID: 0000-0003-0868-8308

SAŽETAK

ključne riječi: digitalni tragovi, osobine ličnosti, psihologija ličnosti, psihometrija, veliki podaci, Velikih pet

Povećanim korištenjem interneta i društvenih mreža, korisnici za sobom ostavljaju sve više različitih podataka koji odražavaju njihova ponašanja, misli i osjećaje, a nazivaju se digitalnim tragovima. Digitalni se tragovi upotrebljavaju u brojnim područjima kako bi se mogla predvidjeti ponašanja korisnika i time poboljšati njihova iskustva korištenja digitalnih usluga i platformi. Nedavno je taj oblik velikih podataka pronašao svoju primjenu i u psihologiji ličnosti što predstavlja veliku promjenu u odnosu na tradicionalne načine prikupljanja podataka u tom području poput samoprocjene i procjena drugih na upitnicima ličnosti. Osim očitih prednosti u vidu lake dostupnosti ogromne količine podataka dobivene nenametljivim putem, korištenje velikih podataka predstavlja i nov izazov za psihologiju kao znanost. Izazovi se prvenstveno tiču psihometrijskih karakteristika velikih podataka kao mjere ličnosti, koje su u dosadašnjim istraživanjima stavljene u drugi plan zbog toga što su istraživači uglavnom radili na maksimalizaciji točnosti predviđanja (samo)procjena ličnosti na temelju digitalnih tragova, a manje na provjeri pouzdanosti, valjanosti ili generalizabilnosti nalaza. Također, neki znanstvenici smatraju da bi se veliki podaci mogli koristiti i za unapređenje teorija ličnosti, no istraživanja tek trebaju krenuti u tom smjeru. Iako su digitalni tragovi vrijedan izvor podataka, važno ih je u istraživanjima upotrebljavati uz pridržavanje etičkih načela.

ABSTRACT

key words: big data, Big Five, digital traces, personality psychology, personality traits, psychometrics

Due to increased internet and social media usage, users are leaving behind more and more different types of data called digital traces, which reflect their behaviour, thoughts, and feelings. Digital traces are being used in numerous fields to predict users' behaviour and improve digital platforms and services. Recently, big data has been used in personality psychology as well, representing a major change in comparison to traditional assessment methods such as self- and informant-report scores on personality questionnaires. Besides some advantages such as being unobtrusive and highly available, using big data in psychology is a challenge. Since the majority of research has been focused mostly on maximizing the accuracy of self- and informant-report scores prediction based on digital traces, there is a severe lack of reliability, validity, and generalizability testing, so the biggest challenge of big data in psychology is psychometric evaluation. Additionally, some researchers find that big data could be used to advance personality theory, but so far, that kind of research has been neglected. Although digital traces are a very valuable source of information, it is important to follow all ethical guidelines and principles when using them.

UVOD

Prema podacima iz 2020. godine, više od četiri milijarde ljudi, odnosno polovina svjetske populacije, koristi društvene mreže (Statista, 2020). Pri tome prosječan korisnik ima osam korisničkih računa na različitim društvenim mrežama te na njima dnevno provodi oko dva i pol sata (DataReportal, 2020). Za to vrijeme, svaki korisnik za sobom ostavlja razne oblike digitalnih tragova – zapisa o svojem ponašanju poput *likeova*, riječi korištenih u objavama i komunikaciji, geografske lokacije, fotografija i slično (Bleidorn i Hopwood, 2018). Osim podataka s društvenih mreža, u digitalne tragove spadaju i drugi zapisi koji su posljedica korištenja *smartphonea*, poput onih o telefonskim pozivima, online transakcijama ili fizičkoj aktivnosti koju bilježe *fitness* aplikacije (Kosinski i sur., 2016). Navedeni digitalni tragovi samo su neki od primjera tzv. velikih podataka koji su u posljednje vrijeme privukli pažnju kod istraživača u području društvenih znanosti pa tako i u psihologiji. Koncept velikih podataka (engl. *big data*) odnosi se na prikupljanje i obradu velikih količina podataka čija se analiza vrši raznim metodama umjetne inteligencije. Jedna od najčešće primijenjenih tehnika, kako općenito tako i u psihologiji ličnosti, jest strojno učenje (engl. *machine learning*) koje se zasniva na tome da se algoritmi „uče“, tj. treniraju na velikim količinama podataka kako bi što bolje mogli predviđati određene ishode (Bleidorn i Hopwood, 2018). Za razliku od uobičajenih ograničenja u psihologijskim istraživanjima poput premalih uzoraka ili pitanja iskrenosti sudionika, veliki podaci, osim što rješavaju problem veličine uzorka, pružaju uvid u misli, emocije i interakcije pojedinaca u *online* okruženju (Azucar i sur., 2018). Ovaj novi način prikupljanja podataka sve se češće upotrebljava u psihologiji ličnosti. S obzirom na to da psihološke osobine utječu na način na koji pojedinac koristi digitalne usluge, provedena su brojna istraživanja koja provjeravaju u kojoj se mjeri, na temelju digitalnih tragova, mogu predviđati osobine ličnosti pojedinca. Dosadašnja istraživanja uglavnom su procjenjivala mogućnost predikcije Velikih pet osobina – ekstraverzije, neuroticizma, ugodnosti, savjesnosti i otvorenosti, najčešće pomoću modela strojnog učenja (Bleidorn i sur., 2017). U ovom će se radu usporediti tradicionalni pristup istraživanju ličnosti i pristup s velikim podacima te će se navesti prednosti i nedostaci novog pristupa iz psihometrijske perspektive. Također, pružit će se uvid u dosadašnja istraživanja, kao i bitne napomene pri korištenju velikih podataka u društvenim znanostima.

PRIKUPLJANJE PODATAKA: TRADICIONALNI PRISTUP VS. VELIKI PODACI

Donedavno su u psihologiji ličnosti glavni izvori podataka bili rezultati na upitnicima ličnosti najčešće dobiveni samoprocjenom, ali i procjenama drugih (Boyd i sur., 2020). Osim vremenske zahtjevnosti i nepraktičnosti ove metode u usporedbi s novijima (Farnadi i sur., 2016), prisutni su i problemi objektivnosti te valjanosti. Sudionici prilikom samoprocjene mogu svjesno ili nesvjesno davati odgovore koji bi se razlikovali od nekih objektivnijih mjera ponašanja (poput opažanja ponašanja), a i sami upitnici nisu uvijek sadržajno valjani (Boyd i sur., 2020). U prošlom je stoljeću

metoda samoprocjene bila jednostavnija od drugih dostupnih objektivnijih metoda pa je zato i bila toliko često korištena. Međutim, s razvojem tehnologije i pojavom koncepta velikih podataka, prikupljanje bihevioralnih mjera postalo je relativno jednostavno, a istovremeno manje podložno gore navedenim problemima (Boyd i sur., 2020). Osim što je nenametljivo i time manje pod utjecajem socijalne poželjnosti (Azucar i sur., 2018), prikupljanje velikih podataka obuhvaća jako velike uzorke, kao i širok spektar ponašanja koja mogu biti indikatori ličnosti. Usporedbom uspješnosti tradicionalnih i suvremenih metoda u predikciji ličnosti bavili su se Youyou i suradnici (2015). U svojem su istraživanju ustanovili da su *Facebook likeovi* (stranice koje su korisnici *likeali* na *Facebooku*) bolji prediktor ($r = .56$) rezultata na upitniku ličnosti dobivenih samoprocjenom, nego što su to procjene bliskih ljudi ($r = .49$), a ta se prednost povećava s povećanjem dostupnog broja *likeova*. Osim toga, kad su kriterijske varijable činili različiti životni ishodi, Facebook likeovi predviđali su neke od ishoda (npr. aktivnost na društvenim mrežama, korištenje psihoaktivnih tvari) točnije čak i od samoprocjene ličnosti. Ovaj je nalaz posebno značajan s obzirom na to da su životni ishodi dobiveni samoprocjenom pa očekivano više koreliraju sa samoprocjenom ličnosti. Iz svega navedenog može se zaključiti da veliki podaci imaju brojne prednosti kojima se mogu nadomjestiti nedostaci tradicionalnih metoda u psihologiji ličnosti, no neizbježno se nameću pitanja što je točno zahvaćeno digitalnim tragovima, odnosno što je to što korelira s rezultatima u testovima ličnosti te radi li se nužno o osobinama ličnosti ili, pak, o nekim drugim konstruktima.

VELIKI PODACI: PSIHOMETRIJSKA PERSPEKTIVA

Pouzdanost. Kako bi se utvrdilo mogu li digitalni tragovi biti adekvatan indikator osobina ličnosti, potrebno je ispitati njihovu pouzdanost, valjanost i generalizabilnost. Prvo je nužno provjeriti pouzdanost, jer tek kad se ustanovi da je riječ o stabilnoj mjeri, ima smisla pitati se koji se konstrukti nalaze u njenoj podlozi. Budući da je kod primjene strojnog učenja cilj maksimalizirati točnost predviđanja pa u ulazne podatke spadaju svi oni koji povećavaju točnost neovisno o njihovom značenju, ne očekuje se da su različiti podaci međusobno visoko povezani. Prema tome, procjena pouzdanosti metodom unutarnje konzistentnosti nije adekvatna za modele strojnog učenja. S druge strane, adekvatnom se pokazala procjena pouzdanosti velikih podataka test-retest metodom (Bleidorn i Hopwood, 2018). Park i suradnici (2015) proveli su jedno od rijetkih istraživanja na tekstualnim podacima u kojima se provjeravala pouzdanost predikcija (Tay i sur., 2020). Povezanost riječi korištenih u Facebook objavama i rezultata na upitniku ličnosti provjeravali su test-retest metodom u četiri vremenske točke pri čemu su se za svaku uzimale riječi korištene u intervalu od šest mjeseci. Prosječna pouzdanost za svih pet osobina ličnosti u svim vremenskim točkama iznosi .70, a raspon se kreće od .62 za neuroticizam do .76 za savjesnost. Što se tiče tradicionalnih metoda, test-retest pouzdanost na upitnicima Velikih pet osobina ličnosti kreće se između .65 i .85 (Park i sur., 2015). Dobiveni su rezultati značajni jer se pojam ličnosti odnosi na relativno trajne obrasce doživljavanja i ponašanja pa je barem umjerena pou-

zdanost nužan preduvjet za razmatranje konstruktne valjanosti. Ipak, radi se o jednoj procjeni pouzdanosti velikih podataka u predviđanju ličnosti i to na temelju tekstualne analize Facebook objava, tako da se iz ovog istraživanja ne može zaključivati o pouzdanosti drugih tipova velikih podataka. Također, Tay i suradnici (2020) za buduća istraživanja predlažu procjenjivanje pouzdanosti metodom slaganja među algoritmima za predviđanje ličnosti, prema analogiji sa slaganjem među procjenjivačima kao klasičnom mjerom pouzdanosti. Tako bi istraživači mogli jasnije razaznati u kojoj su mjeri dobivene predikcije specifične za određeni prediktivni model, odnosno neovisne o modelu koji se koristi (Tay i sur., 2020).

Valjanost. Većina istraživanja provedenih na velikim podacima uglavnom je nastojala maksimalizirati učinkovitost modela strojnog učenja odnosno točnost predviđanja osobina ličnosti (Bleidorn i sur., 2017), zane-marivši pritom valjanost digitalnih tragova kao mjere ličnosti. Kako su kriterijske varijable najčešće bile samoprocjene na upitnicima ličnosti, dosadašnja su istraživanja zapravo dokazivala samo konvergentnu valjanost (Park i sur., 2015), dok su ostali aspekti konstruktne valjanosti slabo istraživani. U svrhu određivanja veličine povezanosti digitalnih tragova i samoprocjena osobina ličnosti provedene su dvije meta-analize prema kojima se korelacije kreću u rasponu od .29 (ugodnost) do .40 (ekstraverzija) za pojedinačne osobine ličnosti (Azucar i sur., 2018), a za ličnost općenito korelacija iznosi .34 (Settanni i sur., 2018). Toliki koeficijenti korelacije odgovaraju uobičajenim veličinama povezanosti između ličnosti i ponašanja (.30 - .40) pa se može zaključiti kako su digitalni tragovi pouzdan izvor informacija za predikciju osobina ličnosti (Azucar i sur., 2018). Izazov kod ovakvog istraživačkog pristupa čini to što se ljudske procjene vlastite ili tuđe ličnosti, koje sadrže pogrešku mjerenja, izjednačavaju s istinom (Farnadi i sur., 2016). Drugim riječima, zadatak prediktivnih modela je da imitiraju nesavršene procjene koje ljudi vrše na temelju ograničenih informacija (Tay i sur., 2020). Iako se taj problem može ublažiti povećanjem točnosti predviđanja tako da se zahvati širi raspon podataka, primjerice s različitim društvenih mreža (Farnadi i sur., 2016), zbog ljudskih pristranosti, pogrešaka i drugih manjkavosti pri procjeni, postavlja se pitanje kako validirati digitalne tragove kao mjeru ličnosti bez referiranja na rezultate upitnika (Tay i sur., 2020). Što se tiče diskriminativne valjanosti, korelacije među osobinama ličnosti značajno su više kod digitalnih tragova nego kod samoprocjena, što ukazuje na manju diskriminativnu valjanost velikih podataka. Općenito je diskriminativna valjanost dobivena na velikim podacima uglavnom loša, uz iznimku istraživanja Parka i suradnika (2015) koji su dobili relativno niske korelacije među osobinama ličnosti (Tay i sur., 2020). Razlog lošoj diskriminativnoj valjanosti je to što se za maksimalizaciju konvergentne valjanosti koriste svi podaci koji je imalo povećavaju, a pritom je velika mogućnost da neki od njih smanjuju diskriminativnu valjanost (Bleidorn i Hopwood, 2018). Kako bi se umanjio taj nedostatak, potrebno je u prediktivne modele uvrštavati prediktore koji istovremeno povećavaju konvergentnu, a ne umanjuju diskriminativnu valjanost (Tay i sur., 2020). Kod ispitivanja kriterijske valjanosti dobiveni su smjerom i veličinom slični obrasci povezanosti s vanjskim varijablama za samoprocjene ličnosti i za tekstualne digitalne tragove, ali su korelacije uglavnom niže za digitalne tragove zbog toga što su podaci vanjskih varijabli, poput

zadovoljstva životom i samomotrenja, također dobiveni samoprocjenom (Park i sur., 2015). Istraživanje mogućnosti predviđanja životnih ishoda i drugih vanjskih varijabli na temelju digitalnih tragova poželjno je kako bi se utvrdilo predviđaju li ih oni bolje od tradicionalnih mjera ličnosti te kako bi se ispitala stabilnost tih predviđanja (Tay i sur., 2020). Osim toga, kriterijska valjanost predstavlja pragmatički aspekt pa su daljnja istraživanja povezanosti s vanjskim varijablama korisna zbog eventualne primjenjivosti nalaza dobivenih na velikim podacima. Sadržajna valjanost digitalnih tragova najmanje je istraživana zbog same prirode algoritama koji su često teško interpretabilni te se nazivaju „crnim kutijama“ (Alexander i sur., 2020). Tako su, primjerice, neki nalazi intuitivni i u skladu s dosadašnjim spoznajama, poput povezanosti *likeanja* stranice *Cheerleading* na *Facebooku* i visoke ekstraverzije, dok su neki drugi visoko prediktivni *likeovi* (npr. *Getting money* za niski neuroticizam) teže razumljivi (Bleidorn i sur., 2017). Ipak, postoje i transparentniji algoritmi čiji je način rada jasniji, ali kod njih je često slučaj da imaju slabiju sposobnost predviđanja tako da se među modelima strojnog učenja često treba birati između predviđanja i razumijevanja (Alexander i sur., 2020). Kako je bez adekvatne sadržajne valjanosti teško govoriti o konstruktnoj, jer je moguće da je u nastojanju povećanja povezanosti varijabli zahvaćena nezanemariva količina konstruktno irelevantne varijance, Bleidorn i Hopwood (2018) predlažu uvođenje procjena stručnjaka koji bi određivali koji su prediktori u skladu s teorijom, odnosno na kojim se podacima mogu trenirati modeli za predviđanje ličnosti. S druge strane, Stachl i suradnici (2020b) smatraju kako prediktori i ne moraju biti interpretabilni u okviru teorije, već da se radi o osobitosti modela strojnog učenja koji pružaju uvid u realističniju strukturu ličnosti koja je kompleksnija od teorijskih postavki ličnosti zbog čega ne preporučuju pretjerano oslanjanje na tradicionalni pristup sadržajnoj valjanosti. Zbog brojnih aspekata valjanosti, kompleksnosti primjene psihometrijskih načela na velike podatke te nedostatka istraživanja koja se bave razumijevanjem ličnosti na temelju velikih podataka, stupanj valjanosti digitalnih tragova kao indikatora ličnosti još nije moguće u potpunosti razumjeti i odrediti.

Generalizabilnost. U istraživanjima na području psihologije ličnosti, različiti tipovi velikih podataka prikupljaju se s društvenih mreža koje se razlikuju prema glavnim obilježjima te načinu komunikacije pa tako i ponašanjima koja se na njima iskazuju. Drugim riječima, karakteristike pojedine platforme poput tipa sadržaja koji se objavljuje (npr. tekst na *Twitteru*, slike na *Instagramu*, videozapisi na *YouTubeu*) i informacija koje se dijele (npr. demografski podaci i vremenska crta događaja na *Facebooku*) pružaju različite kontekste za ekspresiju osobina ličnosti (Tay i sur., 2020). Shodno tome, rezultati istraživanja provedenih na različitim indikatorima ličnosti ili na različitim društvenim mrežama nisu direktno usporedivi, što predstavlja problem za mogućnost generalizacije nalaza (Bleidorn i Hopwood, 2018). Također, u velikoj se mjeri razlikuju i algoritmi korišteni u predviđanju ličnosti zato što su i tipovi podataka u istraživanjima raznoliki, zbog čega se pojavljuje pitanje generalizabilnosti jer nisu svi rezultati istraživanja dobiveni istom metodom (Tay i sur., 2020). Iako su uzorci u istraživanjima povezanosti digitalnih tragova i ličnosti jako veliki, njihova je demografska struktura takva da nisu reprezentativni u

odnosu na populaciju. Kako se radi o korištenju modernih tehnologija, u znatno su većoj mjeri zastupljeni mlađi, urbaniji, obrazovaniji i bogatiji ljudi (Pew Research Center, 2018) jer se dostupnost i korištenje interneta i *smartphoneova* razlikuju među kulturama i dobnim skupinama. S obzirom na sve navedeno, rezultati dosadašnjih istraživanja mogu biti korisni za primjenu u sličnim okolnostima u kojima su dobiveni ili pak, za generiranje novih istraživačkih hipoteza, ali nije ih moguće smatrati u potpunosti važećima na općoj razini.

DOSADAŠNJA ISTRAŽIVANJA LIČNOSTI NA VELIKIM PODACIMA

Većina se istraživanja u ovom području bavila samom povezanošću osobina ličnosti i određenih digitalnih tragova, dok onih usmjerenih na kauzalnost, odnosno na razumijevanje i pružanje teoretskog objašnjenja povezanosti te unapređenje teorija ličnosti, uvelike nedostaje (Settanni i sur., 2018). Bleidorn i Hopwood (2018) sistematizirali su dotadašnja istraživanja i podijelili ih na tri generacije. U prvu generaciju spadaju istraživanja s početka prošlog desetljeća koja su uglavnom provodili znanstvenici iz područja računalstva, a specifična su po tome što su uvela velike podatke u istraživanja ličnosti. Provođena su tako što bi se prikupile mjere ličnosti sudionika nekim od upitnika Velikih pet osobina ličnosti te bi se analiziralo koji su podaci s društvenih mreža (uglavnom *Facebook* i *Twitter*) povezani barem s jednom od tih osobina (Bleidorn i Hopwood, 2018). U istraživanju na 167 korisnika *Facebooka*, analizom riječi korištenih u porukama i objavama, ekstrahirane su varijable koje su bile značajno povezane s rezultatima na upitnicima, nakon čega su te varijable korištene za predviđanje osobina ličnosti pojedinca. Rezultati su pokazali da su predikcije varijabli bolje od slučaja, a najtočnije je predviđena otvorenost (Golbeck i sur., 2011). Broj sudionika u istraživanjima prve generacije relativno je malen te osim zanemarivanja sadržajne valjanosti, ova istraživanja karakterizira i slaba statistička snaga. Međutim, njihova je vrijednost u tome što pružaju dokaz da su digitalni tragovi povezani s osobinama ličnosti korisnika digitalnih usluga (Bleidorn i Hopwood, 2018). U istraživanjima druge generacije korišteni su znatno veći uzorci, a težilo se maksimalizaciji točnosti predviđanja. Glavni razlog povećane statističke snage jest korištenje velikih uzoraka, odnosno podataka prikupljenih putem aplikacije *myPersonality* u sklopu *Facebooka*, gdje su milijuni korisnika, od kojih je dio pristao i na dijeljenje podataka, riješili neke od ponuđenih testova koji su uključivali i procjenu vlastite ličnosti ili ličnosti drugih upitnikom Velikih pet osobina ličnosti (Bleidorn i Hopwood, 2018). Jedno od takvih je i već spomenuto istraživanje Parka i suradnika (2015) koji su na temelju tekstualne analize uspješno (od $r = .35$ za neuroticizam do $r = .43$ za otvorenost) predviđali osobine ličnosti kod korisnika *Facebooka*. Problem istraživanja druge generacije predstavljaju sadržajna i diskriminativna valjanost digitalnih tragova kao mjere ličnosti zbog uvištavanja svega što povećava korelaciju u skup prediktorskih varijabli, ne vodeći pritom računa o sadržaju i odnosu tih varijabli s vanjskim varijablama. Treća generacija uključuje istraživanja koja su koristila postojeće algoritme kako bi se provjerilo jesu li digitalni tragovi bolja mjera ličnosti od tradicio-

nalnih metoda. Jedan je primjer i već spomenuto istraživanje Youyou i suradnika (2015) koji su ustanovili da digitalni tragovi predviđaju ličnost bolje od procjena drugih, dok su u odnosu na samoprocjene nekad bolji, a nekad lošiji. Važno je istaknuti da su modeli strojnog učenja trenirani na samoprocjenama, dok procjene drugih nisu, stoga je upitno koliko su te dvije kategorije usporedive. Također, istraživači se još nisu bavili pitanjima sadržajne valjanosti pa ostaje nejasno koji dio ličnosti digitalni tragovi zahvaćaju (Bleidorn i Hopwood, 2018).

Jedno od rijetkih istraživanja koje se nije bavilo predviđanjem, već je koristilo velike podatke u svrhu unapređenja teorije ličnosti, jest istraživanje Gerlacha i suradnika (2018) u kojem su identificirana četiri tipa ličnosti. Tipološki pristup ličnosti karakterizira to što različiti tipovi trebaju imati različita obilježja na skupu osobina, odnosno to što se ljude dijeli na kategorije ili tipove, dok su prema osobinskom pristupu osobine ličnosti zapravo kontinuumi (Ones i Wiernik, 2019). Osim po svrsi, ovo se istraživanje od nekih ranije spomenutih razlikuje i po tome što velike podatke nisu činili digitalni tragovi, već velik broj rezultata na *online* upitnicima ličnosti. Korištenjem tek razvijene klastering metode, istraživači su na jako velikim uzorcima (četiri uzorka u rasponu od 145 338 do 575 380 sudionika) pronašli četiri tipa ličnosti: prosječan (rezultati na svim osobinama kreću se oko prosjeka), uzoran (nizak rezultat na neuroticizmu, visok na ostalim osobinama), rezerviran (niski rezultati na otvorenosti i neuroticizmu) i usredotočen na sebe (visok rezultat na ekstraverziji, niski rezultati na otvorenosti, ugodnosti i savjesnosti) (Gerlach i sur., 2018). Uz same tipove ličnosti, pronašli su i konzistentne razlike u zastupljenosti pojedinog tipa ličnosti u odnosu na dob i spol. Kako je istraživanje provedeno suvremenim metodama na jako velikim uzorcima i objavljeno u renomiranom časopisu *Nature Human Behaviour*, izazvalo je veliku medijsku pozornost popraćenu brojnim naslovima koji sugeriraju na preokret u psihologiji ličnosti u vidu nove robusne teorije. S druge strane, Ones i Wiernik (2019) ne smatraju da je došlo do otkrića novih tipova ličnosti niti podržavaju tipološki pristup. Naime, budući da su Velikih pet osobina ličnosti u međusobnim korelacijama, očekivano je i poznato da će se neke osobine ličnosti, poput ugodnosti i savjesnosti ($r = .43$), ili ekstraverzije i otvorenosti ($r = .43$), pojavljivati zajedno kao visoke ili niske (Van der Linden i sur., 2010). Također, profili dobiveni istraživanjem ne mogu se interpretirati kao tipovi zbog međusobnih preklapanja odnosno nepostojanja jasne distinkcije među različitim tipovima. Osim toga, velik broj istraživanja pokazao je kako tipovi nisu adekvatna paradigma te da najbolje razumijevanje ličnosti pruža osobinski pristup (Ones i Wiernik, 2019). Na primjer, McCrae i Costa (1989) u svojem su istraživanju pokazali da MBTI (*Myers-Briggs Type Indicator*) tipologija ne identificira kvalitativno različite tipove ličnosti, već da se radi o rezultatima na četiri nezavisne dimenzije kojima nije pokriven faktor neuroticizma. Wilmot i suradnici (2019) su pri analizi TAB (*Type A Behavior*) tipologije dobili da tip A ličnosti zapravo nije tip te da je njegova latentna struktura dimenzionalna. Nadalje, pri pokušaju replikacije nalaza da je samomotrenje kategorija na kojoj pojedinac može biti ili visoko ili nisko, taksometrijskom analizom (metodom kojom se provjerava radi li se o kategoričkom ili dimenzionalnom mod-

elu) pronađeno je da bi se i samomotrenje trebalo konceptualizirati kao dimenzionalno (Wilmot, 2015). Dakle, pri analizi i replikaciji nekih od najpopularnijih nalaza koji podržavaju kategoričku strukturu ličnosti dobiveno je da je u njihovoj podlozi zapravo dimenzionalna struktura.

Unatoč istaknutim nedostacima u istraživanju Gerlacha i suradnika (2018), pohvalno je što se krenulo u smjeru unapređenja razumijevanja ličnosti, čime se možda nagovještava četvrta generacija istraživanja u kojoj fokus više neće biti isključivo na predikciji. Bleidorn i suradnici (2017) predložili su tri područja u kojima bi istraživanja na velikim podacima mogla unaprijediti teoriju ličnosti: razjašnjavanje strukture i sadržaja osobina ličnosti koji se možda razlikuju od onih dobivenih leksičkom hipotezom (stajalište da su individualne razlike u ličnosti ljudi kodirane u jeziku), razumijevanje utjecaja razvojnih procesa na ličnost i određivanje kulturalno specifičnih indikatora ličnosti. Drugim riječima, veliki bi podaci možda mogli dati odgovore na dosad nerazriješena pitanja u području psihologije ličnosti ili pak redefinirati postojeće spoznaje.

PRAKTIČNA PRIMJENA I PITANJE PRIVATNOSTI

Osim spoznajne, istraživanja ličnosti na velikim podacima imaju i praktičnu vrijednost. Mogućnost predviđanja osobina ličnosti na temelju digitalnih tragova može se koristiti za prilagođavanje digitalnih i drugih usluga pojedincu (Kosinski i sur., 2013). Neke od korisnih primjena su poboljšanje korisničkog iskustva na digitalnim uslugama (Farnadi i sur., 2016), kreiranje personaliziranih zdravstvenih intervencija i poruka (budući da osobine ličnosti mogu imati ulogu zaštitnog ili rizičnog faktora) te personalizirani sustav preporuka i oglasa (Azucar i sur., 2018). Međutim, dostupnost i korištenje podataka koji pružaju uvid u privatnost ljudi otvara etička pitanja. Društvene mreže i ostale digitalne platforme omogućuju dijeljenje privatnih informacija u mnogo većem obujmu i s mnogo više detalja nego prije (Kosinski i sur., 2015). Iako se prilikom upotrebe usluga traži pristanak na dijeljenje određenih podataka, korisnicima bi trebalo biti jasnije koji se tipovi podataka i u koje svrhe koriste, kao i da postoji mogućnost prodaje ili dijeljenja tih podataka s trećim stranama. Na primjer, korisnici bi trebali biti svjesni toga da se njihovi podaci mogu koristiti i u svrhu predviđanja nekih karakteristika poput osobina ličnosti ili potrošačkog ponašanja, a ne samo kako bi se pratilo korištenje određene aplikacije (Stachl i sur., 2020a). O zlouporabi podataka i mogućnosti manipuliranja korisnicima sve se više priča nakon slučaja *Facebooka* i *Cambridge Analytica* koji su bez pristanka koristili podatke milijuna korisnika i podijelili ih s trećom stranom u svrhu vođenja političke kampanje. Takvi postupci predstavljaju opasnu prijetnju, ne samo samostalnom donošenju (glasačkih) odluka, već i demokraciji općenito (Boyd i sur., 2020). Ovaj je problem prepoznalo i zakonodavstvo pa je 2016. godine donesena *Opća uredba o zaštiti podataka* (GDPR) na području Europske unije kojom se regulira privatnost i zaštita podataka pojedinca. Sve u svemu, iako su veliki podaci poput digitalnih tragova iznimno vrijedan izvor podataka za akademsku zajednicu, potrebno se prilikom prikupljanja i korištenja podataka pridržavati zakonskih odrednica i etičkih načela vezanih za privatnost korisnika.

ZAKLJUČAK

Razvoj tehnologije i veliki napreci u području računalstva i podatkovne znanosti doveli su do toga da se na temelju digitalnih tragova koje ljudi ostavljaju na društvenim mrežama i mobilnim uređajima mogu predviđati razne karakteristike pojedinca, uključujući i osobine ličnosti. Novi način prikupljanja podataka u posljednjem je desetljeću usmjerio istraživanja u području psihologije ličnosti prema sasvim novom, interdisciplinarnom pristupu koji spaja psihologiju i podatkovnu znanost. U odnosu na tradicionalne metode u psihologiji ličnosti, veliki podaci pokrivaju neke od nedostataka poput premalih uzoraka ili iskrivljavanja istine pri odgovaranju na upitnicima. Dosadašnja istraživanja uglavnom su se bavila predviđanjem osobina ličnosti na temelju digitalnih podataka. Iako su modeli za predviđanje u velikoj mjeri napredovali, još uvijek nije došlo do značajnijih doprinosa teoriji ličnosti pa se očekuje da će se buduća istraživanja više usmjeriti na objašnjavanje, a manje na predviđanje ličnosti. Ipak, dosadašnja su istraživanja pokazala i nedostatke modernog pristupa iz gledišta psihometrije. Naime, iako su modeli za predviđanje ličnosti sve precizniji, i dalje nije jasno što je točno njima zahvaćeno, tim više što se modeli treniraju kako bi predviđali (samo)procjene ljudi koje su same po sebi opterećene pogreškom. Također, osim prednosti korištenja velikih podataka u vidu njihove široke primjene, postoje i nedostaci poput mogućnosti povrede privatnosti. Pitanje etičnosti u društvu se nameće kao sve važnije pa se prema tome i djeluje usvajanjem zakonskih regulativa za zaštitu i privatnost podataka. S obzirom na sve navedeno, može se reći da su veliki podaci pronašli svoju primjenu i u psihologiji ličnosti, ali se barem zasada, ne radi o metodi koja može zamijeniti tradicionalne, već bi se trebale međusobno nadopunjavati. Iako brojni istraživači u ovom pristupu vide velik potencijal za dopunjavanje manjkavosti i razvoj teorija ličnosti, taj potencijal još uvijek nije do kraja iskorišten.

LITERATURA

- Alexander, L., Mulfinger, E. i Oswald, F. L. (2020). Using Big Data and Machine Learning in Personality Measurement: Opportunities and Challenges. *European Journal of Personality*, 34(5), 632-648. <https://doi.org/10.1002/per.2305>
- Azucar, D., Marengo, D. i Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124, 150-159. <https://doi.org/10.1016/j.paid.2017.12.018>
- Bleidorn, W., Hopwood, C. J. i Wright, A. G. C. (2017). Using big data to advance personality theory. *Current Opinion in Behavioral Sciences*, 18, 79-82. <https://doi.org/10.1016/j.cobeha.2017.08.004>
- Bleidorn, W. i Hopwood, C. J. (2018). Using Machine Learning to Advance Personality Assessment and Theory. *Personality and Social Psychology Review*, 23(2), 190-203. <https://doi.org/10.1177/1088868318772990>
- Boyd, R. L., Pasca, P. i Lanning, K. (2020). The Personality Panorama: Conceptualizing Personality Through Big Behavioural Data. *European Journal of Personality*, 34(5), 599-612. <https://doi.org/10.1002/per.2254>
- DataReportal. (3. siječnja 2020). Global Social Media Overview. <https://datareportal.com/social-media-users>
- Gerlach, M., Farb, B., Revelle, W. i Nunes Amaral, L. A. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour*, 2, 735-742. <https://doi.org/10.1038/s41562-018-0419-z>
- Golbeck, J., Robles, C. i Turner, K. (2011). Predicting personality with social media. *Proceedings from CHI EA '11: Annual conference extended abstracts on Human factors in computing systems*, 253-262. <https://doi.org/10.1145/1979742.1979614>
- Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stilwell, D., Davalos, S., Moens, M. F. i De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, 26(2-3), 109-142. <https://doi.org/10.1007/s11257-016-9171-0>
- Kosinski, M., Stillwell, D. i Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences (PNAS)*, 110(15), 5802-5805. <https://doi.org/10.1073/pnas.1218772110>
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V. i Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543-556. <https://doi.org/10.1037/a0039210>
- Kosinski, M., Wang, Y., Lakkaraju, H. i Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, 21(4), 493-506. <https://doi.org/10.1037/met0000105>
- McCrae, R. R. i Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality. *Journal of Personality*, 57(1), 17-40. <https://doi.org/10.1111/j.1467-6494.1989.tb00759.x>
- Ones, D. S. i Wiernik, B. M. (2019). On "new" personality types: An industrial, work, and organizational psychology perspective. *The Industrial-Organizational Psychologist*, 56(3).
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H. i Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934-952. <https://doi.org/10.1037/pspp0000020>
- Pew Research Center. (12. veljače 2021). Social Media Use in 2018. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/02/PI_2018.03.01_Social-Media_FINAL.pdf
- Settanni, M., Azucar, D. i Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking*, 21(4), 217-228. <https://doi.org/10.1089/cyber.2017.0384>
- Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bissch, B. i Bühner, M. (2020a). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences (PNAS)*, 117(30), 17680-17687. <https://doi.org/10.1073/pnas.1920484117>
- Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D. i Bühner, M. (2020b). Personality Research and Assessment in the Era of Machine Learning. *European Journal of Personality*, 34(5), 613-631. <https://doi.org/10.1002/per.2257>
- Statista. (3. siječnja 2020). Global digital population as of October 2020. <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- Tay, L., Woo, S. E., Hickman, L. i Saef, R. M. (2020). Psychometric and Validity Issues in Machine Learning Approaches to Personality Assessment: A Focus on Social Media Text Mining. *European Journal of Personality*, 34(5), 826-844. <https://doi.org/10.1002/per.2290>
- Van der Linden, D., Nijenhuis, J. i Bakker, A. B. (2010). The

General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, 44(3), 315-327. <https://doi.org/10.1016/j.jrp.2010.03.003>

Wilmot, M. P. (2015). A contemporary taxometric analysis of the latent structure of self-monitoring. *Psychological Assessment*, 27(2), 353-364. <https://doi.org/10.1037/pas0000030>

Wilmot, M. P., Haslam, N., Tian, J. i Ones, D. S. (2019). Direct and conceptual replications of the taxometric analysis of type a behavior. *Journal of Personality and Social Psychology*, 116(3), 12-26. <https://doi.org/10.1037/pspp0000195>

Youyou, W., Kosinski, M. i Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences (PNAS)*, 112(4), 1036-1040. <https://doi.org/10.1073/pnas.1418680112>