Primjena umjetne inteligencije putem virtualnih glasovnih asistenata

Jurković, Matko

Undergraduate thesis / Završni rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:318177

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Filozofski fakultet
Sveučilišta u Zagrebu

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
Ak. god. 2018./ 2019.

Matko Jurković

Primjena umjetne inteligencije putem virtualnih glasovnih asistenata

Završni rad

Mentor: dr. sc. Ivan Dunđer

Zagreb, lipanj 2019.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

_____

(potpis)

Sadržaj

Sadržaj iv

1. Uvod 1

2. Općenito o umjetnoj inteligenciji 2

3. Chatbotovi 6

4. Općenito o virtualnim glasovnim asistentima 8

5. Prepoznavanje govora i obrada jezika u virtualnim glasovnim asistentima 12

6. Primjeri virtualnih asistenata 14

6.1. Siri 14

6.2. Google Assistant 17

6.3. Cortana, Alexa i Bixby 19

7. Zaključak 21

8. Literatura 22

Sažetak 25

Summary 26

1. Uvod

Virtualni glasovni asistenti danas su posvuda oko nas. Imamo ih na svojim pametnim telefonima, računalima, televizorima, satovima, automobilima kao i uskoro na brojnim drugim uređajima. Kao takvi, oni su otvorili posve nove mogućnosti korištenja navedenih uređaja koje su prije toga bile nezamislive. Bitno je istaknuti kako ne postoji jasno suglasje oko samog naziva ovog tipa aplikacija; tako na primjer, možemo pronaći nazive poput virtualni asistent, AI (eng. *artificial intelligence*) asistent, digitalni asistent, glasovni asistent, osobni asistent, kao i razne kombinacije navedenih (primjerice, virtualni glasovni asistent). Potrebno je biti oprezan pri korištenju nekih od navedenih naziva jer oni mogu imati posve druga značenja. Virtualni asistent može označavati i profesiju, a osobni asistent može označavati i danas već zastarjele dlanovnike (PDAs ili eng. *personal virtual assistants*). Iz tog razloga, u nastavku ovoga rada koristit ćemo izraz virtualni glasovni asistent.

U ovom radu bit će objašnjene dvije discipline koje stoje u pozadini virtualnih glasovnih asistenata, a to su umjetna inteligencija i obrada prirodnog jezika. Danas računalno prepoznavanje ljudskog govora olako shvaćamo, a nismo svjesni koliko je bio dug i kompleksan proces postizanja istog, a koji je temelj funkcioniranja virtualnih glasovnih asistenata. Nakon što se izvrši prepoznavanje izgovorenog upita od strane korisnika, na scenu stupa umjetna inteligencija, obavljajući svoj dio posla. Njezin zadatak je izvršavanje postavljene naredbe ili kreiranje odgovora na korisnikov upit.

U drugom poglavlju bit će objašnjene osnove razvoja kao i područja umjetne inteligencije. U trećem poglavlju ukratko će se objasniti *chatbotovi*, a dan je i njihov kratki povijesni pregled. Četvrto poglavlje je rezervirano za opisivanje osnovnih funkcionalnosti virtualnih glasovnih asistenata kao i njihove zajedničke karakteristike. Peto poglavlje daje kratak uvid u razvoj tehnologije prepoznavanja govora i obrade prirodnog jezika. Šesto poglavlje je određeno za opisivanje danas najraširenijih virtualnih glasovnih asistenta te za specifičnosti svakoga od njih. Cilj ovoga rada je prikazati funkcioniranje i moguće namjene virtualnih glasovnih asistenata, kao i opis tehnologija koje stoje u njihovoj pozadini.

2. Općenito o umjetnoj inteligenciji

Umjetna inteligencija kao fenomen prisutna je već desetljećima, iako u javnom diskursu vidimo njenu značajnu prisutnost tek zadnjih par godina. Razlog tome je njezin rapidni razvoj i širenje kao i njezino promoviranje od strane mnogih tvrtki. Polje umjetne inteligencije ima za cilj pokušaj razumijevanja inteligentnih entiteta (Russell i Norvig, 1995: 3). Stoga, još jedna svrha razvijanja iste može biti razumijevanje nas ljudi kao takvih. Za razliku od filozofije i psihologije, koje se bave samo izučavanjem inteligencije, znanost o umjetnoj inteligenciji teži i izgradnji inteligentnih entiteta kao i razumijevanju istih (Russell i Norvig, 1995:4). Mnogi autori su ponudili svoje definicije umjetne inteligencije koje stavljaju naglasak na njene razne aspekte.

Russell i Norvig (1995:5) definicije umjetne inteligencije dijele u četiri osnovne skupine. Prva skupina definicija ih objašnjava kao sustave koji razmišljaju kao ljudi. Za primjer ovdje imamo definiciju Richarda Bellmana koji umjetnu inteligenciju definira kao automatizaciju aktivnosti koje mi ljudi povezujemo s ljudskim razmišljanjem, aktivnosti kao što su donošenje odluka, rješavanje problema i učenje. Druga skupina ih objašnjava kao sustave koji se ponašaju poput ljudi. Ovdje za primjer Kurzweil umjetnu inteligenciju definira kao umijeće stvaranja strojeva koji obavljaju funkcije koje zahtijevaju inteligenciju kada su obavljani od strane ljudi. Treća skupina definicija umjetne inteligencije ih definira kao sustave koji razmišljaju racionalno. Patrick Winston ovdje nudi definiciju umjetne inteligencije kao studije izračuna koji omogućuju percepciju, razum i djelovanje. Četvrta skupina definicija ih definira kao sustave koji se ponašaju racionalno. Robert J. Schalkoff ovaj tip umjetne inteligencije definira kao polje istraživanja koje nastoji objasniti i emulirati inteligentno ponašanje vezano za računalne procese (Russell i Norvig, 1995)[1].

S obzirom na to da je umjetna inteligencija simulacija ljudske percepcije i djelovanja, njezina podjela na grane izravno je povezana s tim. Ljudi mogu govoriti i slušati kako bi ostvarili komunikaciju putem jezika. Područje umjetne inteligencije koje se bavi ovim zadatkom naziva se prepoznavanje govora (eng. *speech recognition*). Većina posla obavljena od strane sustava za prepoznavanje govora je statistički utemeljena te se zbog toga ponekad veže uz pojam statističkog učenja (eng. *statistical learning*). Ljudi također mogu pisati i čitati tekst na

[1] Sve četiri definicije preuzete su iz knjige Russella i Norviga pod nazivom "Artificial Intelligence: A Modern Approach".

određenom jeziku, stoga područje obrade prirodnog jezika (eng. *natural language processing*) ima za svrhu osposobiti računala razumijevanju ljudskog prirodnog jezika (Ramesh, 2017).

Ljudi zahvaljujući osjetilu vida mogu vidjeti i u svom mozgu obraditi ono što vide. Za obavljanje toga od strane računala brine se računalni vid (eng. *computer vision*). Računalni vid spada pod još jednu granu umjetne inteligencije koja se veže uz pojam simboličkog učenja (eng. *symbolic learning*). Ljudi su svjesni svoje okoline, tj. prostora koji ih okružuje te se mogu fluidno kretati kroz njega. Pod simboličko učenje spada i grana nazvana „robotika“ koja za cilj ima osposobljavanje računala za kretanje i percepciju prostora (Ramesh, 2017).

Čovjek ima značajnu sposobnost uvidjeti i razumjeti ponavljajuće obrasce, kao što je grupiranje objekata s nekim zajedničkim svojstvima. Prepoznavanje obrazaca (eng. *pattern recognition*) je grana umjetne inteligencije koja ovu sposobnost daje računalima. Danas su računala u ovome mnogo uspješnija od ljudi, ponajprije zato što samostalno mogu obraditi ogromne količine podataka i njihova svojstva (Ramesh, 2017).

Ovdje dolazimo do strojnog učenja (eng. *machine learning*) koje računalima omogućava samostalno učenje kako izvršiti određeni zadatak obrađujući velike količine podataka koji upućuju na taj zadatak. Na strojno učenje nadovezuju se neuralne mreže (eng. *neural networks*) koje repliciraju strukturu neurona ljudskog mozga kako bi računalima približile ljudski način razmišljanja. Neuralne mreže mogu biti izrazito kompleksne te se kao takve koriste za najzahtjevnije zadaće koje računala obavljaju, što se još naziva i duboko učenje (eng. *deep learning*) (Ramesh, 2017).

1950. godine Alan Turing (1912.-1954.) objavljuje svoj rad *Computing Machinery and Intelligence*, objašnjavajući u njemu temelje procjene inteligencije strojeva, danas znane pod nazivom Turingov test. Test funkcionira tako da jedna osoba u ulozi procjenitelja ima zadatak voditi razgovor s drugom osobom koja je u drugoj prostoriji kao i računalom koje je u trećoj prostoriji bez da zna tko je osoba, a tko računalo. Razgovor se odvija na prirodnom jeziku, ali u tekstualnom obliku putem zasebnog računala. U slučaju da procjenitelj ne uspije uočiti razliku između osobe i računala, smatra se da je spomenuto računalo (tj. njegov softver koji je imao tu ulogu) prošlo Turingov test. Računalo bi trebalo posjedovati mogućnosti obrade prirodnog jezika, reprezentacije znanja, automatskog rasuđivanja kao i strojnog učenja kako bi uspjelo položiti Turingov test. Turingov test ne uključuje direktnu fizičku interakciju između osobe i računala jer se smatra da fizička simulacija osobe nije nužna za simulaciju ljudske inteligencije. Unatoč tome, postoji i neslužbena inačica Turingovog testa zvana Totalni

Turingov test koja bi trebala uključivati i vid od strane računala, kao i robotiku za svrhu kinetičke interakcije kako bi vjerno simulirala sve vidljive aspekte ljudskog bića (Russell i Norvig, 1995:6).

Kada govorimo o povijesti razvoja umjetne inteligencije svakako je nužno spomenuti rad Marvina Minskyja i Johna McCarthyja. Marvin Minsky (1927.-2016.) bio je američki matematičar, kognitivni znanstvenik te računalni inženjer, također s McCarthyjem zvan ocem umjetne inteligencije. Minsky je vjerovao kako ljudski um u svojoj suštini nije ništa drugačiji od računala te se odlučio fokusirati na inženjerstvo inteligentnih strojeva, najprije u Lincoln laboratoriju, kasnije kao profesor na MIT-u (*Massachusetts Institute of Technology*) gdje je osnovao Laboratorij za razvoj umjetne inteligencije s kolegom Johnom McCarthyjem. Objavio je mnoge radove vezane uz teoriju računanja tako postavivši temelje za razvoj znanosti o umjetnoj inteligenciji te je također pridonio razvoju simboličkog učenja kao grane umjetne inteligencije koja uključuje konceptualne reprezentacije logike i mišljenja visoke razine. Za razliku od mnogih stavova tog vremena koji su upozoravali kako umjetna inteligencija predstavlja opasnost za čovječanstvo, Minsky je na umjetnu inteligenciju gledao blagonaklono vjerujući kako bi ona mogla riješiti neke od najvećih problema s kojima se susreće čovječanstvo (Knight, 2016). Njegov kolega i suradnik John McCarthy također je postavio temelje razvoju umjetne inteligencije. Osmislio je naziv umjetna inteligencija. Zadužen je i za razvoj programskog jezika Lisp koji je postao standardni jezik za razvoj umjetne inteligencije te se koristi čak i danas, pridonijevši tako razvoju tehnologije prepoznavanja govora. Iako su nakon nekog vremena prekinuli suradnju jer su im se perspektive o umjetnoj inteligenciji počele razlikovati, ostali su zapamćeni kao očevi umjetne inteligencije jer su postavili temelje za njen kasniji razvoj (Childs, 2011).

Zanimljiv je i misaoni eksperiment „kineske sobe“ koji je 1980. godine postavio američki filozof John R. Searle (r. 1932.) u svom djelu Minds, Brains and Programs. Za izvedbu ovog eksperimenta potrebno nam je računalo koje zna kineski jezik, tako da uspijeva primiti rečenicu na kineskom jeziku kao ulaznu jedinicu, obraditi ju te složiti odgovor na nju na kineskom jeziku. U slučaju da ovo računalo uspije uvjeriti izvornog govornika kineskog jezika da je ono osoba koja govori kineski jezik, mogli bismo reći da je spomenuto računalo prošlo Turingov test. John Searle ovdje izražava dvojbu razumije li računalo uistinu kineski jezik, ili samo simulira razumijevanje kineskog jezika. Kako bi ovo empirijski provjerio, postavlja sljedeći scenarij. Zamišlja sebe kako sjedi u zatvorenoj sobi držeći knjigu s engleskom verzijom softvera koji omogućuje znanje kineskog jezika prije spomenutom računalu te papir i olovku

kako bi ručno pratio naputke dane od strane prije spomenutog softvera. Izvorni kineski govornik koji se nalazi izvan te sobe bi onda slao papire na kojima su napisane rečenice na kineskom jeziku Searleu koji bi zatim te rečenice obradio prema uputama od strane softvera te nazad poslao repliku na kineskom jeziku na papiru izvornom kineskom govorniku. Searle tvrdi kako je on u ovom slučaju obavio zadatak na identičan način kao i računalo koje je prošlo Turingov test, tj., i on i računalo su pratili upute tako simulirajući znanje kineskog jezika (Searle naglašava da on ne zna kineski jezik). Dakle, ako Searle ne razumije kineski jezik, možemo zaključiti kako ni računalo ne razumije kineski jezik. Searle zaključuje da računalo ne može razumjeti tu konverzaciju jer ne razmišlja te zbog toga ne može ni imati inteligentan um. Ovim je Searle postavio teorijske temelje za glavnu podjelu umjetne inteligencije, onu na snažnu umjetnu inteligenciju i slabu umjetnu inteligenciju.

Snažnu umjetnu inteligenciju Searle definira kao računalo programirano na odgovarajući način kako bi uistinu postalo um, imajući sposobnost razumijevanja te ostala kognitivna stanja. S druge strane, pozicija koja tvrdi kako računala mogu izgledati i djelovati inteligentno, ali bez da imaju sposobnost razumijevanja (posjedovanje uma) nazvana je slabom umjetnom inteligencijom. Takve vrste računala (odnosno softvera) najčešće su specijalizirane za obavljanje određene vrste zadataka s nekog uskog područja (primjerice softver koji igra šah). Esencija ove podjele je pitanje može li računalo doslovno imati um, ili može samo simulirati posjedovanje istog. Ovdje navedena dvojba više je filozofskog karaktera, dok je informatičarima jedini cilj generalno postizanje inteligentnog ponašanja od strane računala (Neapolitan i Jiang, 2018).

Virtualni glasovni asistenti koji su glavna tema ovog rada mogu biti smatrani slabom umjetnom inteligencijom, jer putem algoritama samo simuliraju razumijevanje naredbi, tj. ne razumiju ih same za sebe.

3. Chatbotovi

U ovom poglavlju objasnit će se ukratko povijest i način funkcioniranja *chatbotova,* koji su svojevrsni prethodnici virtualnih glasovnih asistenata kakve poznajemo danas. Technopedia *chatbota* definira kao program pokretan umjetnom inteligencijom koji simulira interaktivan razgovor s čovjekom koristeći unaprijed određene korisničke fraze uz auditivne ili tekstualne signale („What is a Chatbot?“, n.d). Iako po ovoj definiciji virtualni glasovni asistenti pripadaju *chatbotovima*, ovdje ćemo napraviti distinkciju.

Chatbotovi koji će se u ovom poglavlju obrađivati prvenstveno ostvaruju interakciju s korisnikom putem tekstualnog unosa, a ne putem glasa. Oni se često koriste za bazične zadatke u ulozi službe za korisnike i sustava za marketing koji funkcioniraju putem društvenih mreža i *instant messaging* klijenata. Stoga ovdje ne uvrštavamo Siri, Google Assistant, Alexu, Cortanu i ostale velike virtualne glasovne asistente današnjice (iako i oni danas svi nude mogućnost tekstualnog unosa korisničkog upita), već one jednostavnija tipa koji su bili popularni dok virtualni glasovni asistenti kakve danas poznajemo još nisu postali toliko rašireni. Danas se ovakav tip *chatbotova* najčešće koristi kao oblik komunikacije između tvrtki i njihovih korisnika u svrhu informiranja o proizvodima i uslugama kao i rješavanje problema i upita vezano za iste. Njihove prednosti u odnosu na razgovor sa stvarnim ljudima u ovu svrhu je njihova dostupnost bilo kada, instantno odgovaranje na upit te mogućnost jednog *chatbota* da odgovara na upite gotovo neograničenog broja korisnika u isto vrijeme, dok to s ljudskim *chat* agentima nije slučaj (Chatbots In 2019, n.d.).

Prvi poznati *chatbot* bila je Eliza razvijena 1966. godine. Njezin zadatak bio je imitiranje ponašanja psihologa. Kako bi u tome uspjela, koristila je jednostavno poklapanje obrazaca (eng. *pattern matching*) i uglavnom je korisnicima postavljala pitanja. Iako njezina sposobnost održavanja razgovora nije bila na zavidnoj razini, u vrijeme šezdesetih godina prošlog stoljeća bila je dovoljna da stvori osjećaj zbunjenosti kod ljudi koji tada nisu mogli ni zamisliti razgovor s računalom (Bradeško i Mladenić, 2012).

Prvi *chatbot* koji je bio evaluiran koristeći neki vid Turingovog testa bio je Parry 1972. godine. Parry je bio dizajniran kako bi oponašao paranoičnu osobu. Transkripti njegovih izjava te transkripti izjava stvarnih ljudi s dijagnozom paranoje poslani su psiholozima na analizu. Zanimljivo je kako su psiholozi uspjeli razdvojiti Parryja od stvarnih ljudi u samo 48% slučajeva (Bradeško i Mladenić, 2012).

Loebnerova nagrada je godišnja dodjela priznanja najboljim *chatbotovima*, gdje su oni formalno testirani Turingovim testom. Ova dodjela nagrada izazvala je podijeljene stavove oko toga pridonosi li ona uistinu razvoju umjetne inteligencije ili ga zaustavlja. Kritičari ove dodjele nagrada ističu problem razvoja *chatbotova* da jednostavno imitiraju ljudsko ponašanje bez posjedovanja stvarne inteligencije. Jedna od metoda koje neki od njih koriste su i namjerne greške u pisanju te ispravci istih (ovo im je omogućeno jer se njihova demonstracija vrši i uvidom u njihov tekstualni unos). Još jedan problem vezan za ovo natjecanje je poticanje istraživača da rade odvojeno na razvoju umjetne inteligencije, umjesto zajedno te tako dolazi do međusobno nekompatibilnih *chatbotova*. Iako niti jedan *chatbot* na ovom natjecanju do sada nije uspio službeno proći Turingov test, svake godine odabire se pobjednik koji najvjernije simulira komunikaciju s ljudima (Bradeško i Mladenić, 2012).

Tehnologije koje *chatbotovi* koriste kako bi postigli što vjerniju mogućnost vođenja razgovora s ljudima su poklapanje obrazaca (eng. *pattern matching*), parsiranje, korištenje *Markov Chain* modela, ontologije, odnosno semantičke mreže, AIML (*Artificial Intelligence Markup Language*) koji je temeljen na XML-u te ChatScript koji je svojevrsni napredniji nasljednik AIML-a (Bradeško i Mladenić, 2012). *Chatbot* pod imenom Mitsuku, razvijen AIML tehnologijom od strane Stevea Worswicka, četiri je puta osvojio Leobnerovu nagradu (2013., 2016., 2017. i 2018. godine) tako pomevši konkurenciju (Al-Rifaie, n.d.).

4. Općenito o virtualnim glasovnim asistentima

Virtualnog glasovnog asistenta možemo ukratko definirati kao aplikaciju koja razumije glasovne naredbe izrečene na prirodnom jeziku i preko njih obavlja razne zadatke za korisnika (Rouse, 2017). Oni su danas sastavni dio mnogih uređaja koje posjedujemo, a to su pametni telefoni, tableti, kućni zvučnici, pametni satovi, računala, televizori, automobili, a uskoro i mnogih drugih. Najpoznatiji virtualni glasovni asistenti današnjice su Appleova Siri, Googleov Google Assistant, Amazonova Alexa, Microsoftova Cortana i Samsungov Bixby.

U sljedećim poglavljima bit će riječi o specifičnostima svakoga od njih, dok ćemo ovdje istražiti njihove zajedničke karakteristike. Osnovno funkcioniranje virtualnih glasovnih asistenata može se opisati kao pokretanje njihove aplikacije pritiskanjem određene tipke na uređaju ili izgovaranjem glasovne naredbe tipa „*Hey Siri*“ ili „*Okay Google*“, nakon čega korisnik postavlja upit ili zadaje zadatak asistentu, a zatim asistent više ili manje uspješno daje odgovor ili obavlja zadano. Iako svaki virtualni glasovni asistent ima svoje posebne značajke, većina njih funkcionira na sličan način nudeći iste opcije.

Neke od najčešćih zadataka koje virtualni glasovni asistenti izvršavaju su: slanje i čitanje tekstualnih poruka, obavljanje poziva, odgovaranje na jednostavne upite kao što su „*What time is it?, What's the weather forecast for tomorrow?, How many grams does one tonne have?*“ i slično. Virtualni asistenti mogu obaviti i zadatke kao što su postavljanje alarma, unošenje zapisa u aplikaciju kalendara, postavljenje podsjetnika, stvaranje popisa, obavljanje jednostavnih matematičkih izračuna, kontroliranje reprodukcije glazbe na uređaju, upravljanje nekim objektima u stambenom prostoru ako je isti dio sustava pametnog doma i slično. Jedna od svakako najzanimljivijih, ali i za virtualne glasovne asistente najzahtjevnijih radnji su upiti u kojima je potrebno iz internetskih izvora pronaći odgovor i oblikovati ga u smislenu rečenicu. U ovu skupinu pripadaju upiti tipa „*How tall is The Eiffel Tower?, How many atoms are there in the universe?, How many children does Beyoncé Knowles have?*“ i slični. U posljednje vrijeme neki proizvođači virtualnih glasovnih asistenata ubacuju razne zabavne mogućnosti, tako je primjerice moguće zatražiti virtualnog glasovnog asistenta da ispriča šalu, zagonetku, otpjeva pjesmu, recitira poeziju, *beatbox-a* i slično. Korisna opcija je povezivanje s drugim aplikacijama na pametnom telefonu kao što su aplikacije za dijeljenje prijevoza (na primjer Uber i Lyft) ili aplikacijama raznih restorana za naručivanje hrane (na primjer Domino's Pizza). Tako primjerice možemo naručiti Uber vožnju ili određenu vrstu pizze putem virtualnog glasovnog asistenta bez otvaranja aplikacije. Ista stvar je moguća i s kontrolom glazbene

reprodukcije za aplikacije kao što su Spotify ili Pandora. Iako su mnogi virtualni glasovni asistenti od samih početaka bili razvijani za pametne telefone, neki od njih su pomogli popularizirati novi tip multimedijskog uređaja, a to je pametni zvučnik. Ovdje je nužno spomenuti Amazon i njihov virtualni glasovni asistent nazvan Alexa, koja je dostupna na Amazon Echo pametnom zvučniku. Uvidjevši kako su pametni zvučnici postali popularni, Google, Apple i Samsung također izbacuju svoje inačice pametnih zvučnika opremljenih njihovim virtualnim glasovnim asistentima.

Svakako je potrebno spomenuti mogućnosti olakšanog i prilagođenog korištenja tehnologije koje virtualni glasovni asistenti omogućuju osobama s raznim tjelesnim oštećenjima. Ovdje postoji veliki potencijal za pomoć osobama s oštećenjima vida kao i motoričkim oštećenjima. Osobe oštećena vida mogu putem glasovnih naredbi zadanih virtualnom glasovnom asistentu obavljati pozive, slati i čitati SMS poruke, postavljati alarme, kupovati preko interneta kao i mijenjati postavke svog pametnog telefona. Također korisna funkcija osobama s oštećenjima vida je iščitavanje trenutno prikazanog sadržaja na zaslonu pametnog telefona. Virtualni glasovni asistenti na sličan način mogu osobama s motoričkim poteškoćama olakšati korištenje pametnog telefona. Tako je moguće umjesto standardnog upravljanja pametnim telefonom prstima, koristiti mnoge njegove funkcije isključivo glasovnim naredbama (Chan, 2019). Nadalje, osobe s problemima kretanja također mogu imati koristi od virtualnih glasovnih asistenata, jer oni za njih mogu obaviti neke zadatke tipa gašenja svjetla u određenoj prostoriji, zaključavanja vrata, isključivanja ili uključivanja televizora kao i brojne druge interakcije s ostalim kućanskim aparatima. Tako osobe s problemima kretanja ne moraju se bespotrebno kretati stvarajući si opterećenje. Virtualni glasovni asistenti mogu pomoći i ljudima koji boluju od demencije, primjerice pomoću glasovnog podsjetnika mogu dobiti obavijest u vrijeme kada moraju uzeti svoju dnevnu terapiju, nazvati nekoga, otići u kupovinu i ostalo (Medeiros, 2018).

Ovdje postoji i druga strana medalje, a to je nemogućnost virtualnih glasovnih asistenata da jasno razumiju osobe koje imaju oštećenja govora. Ovo može biti problem jer neke bolesti koje otežavaju kretanje i motoriku istovremeno otežavaju i govor (primjerice Parkinsonova bolest, Cerebralna paraliza, ozljede mozga, čak i starija dob), onemogućujući oboljelim osobama gotovo bilo kakvu vrstu interakcije s tehnologijom. Glavni problem nastaje kada virtualni glasovni asistent jednostavno ne uspijeva razumjeti korisnički upit zbog otežanog izgovora uzrokovanog raznim bolestima ili poremećajima. Prilikom razvijanja virtualnih glasovnih asistenata, odnosno njihove sposobnosti razumijevanja ljudskog govora, u većini slučajeva se koristi govor mladih i potpuno zdravih ljudi, tako ograničujući opseg raznolikosti govora koje

asistent na kraju razumije (Corcoran, 2018). Sličan problem koji mnogi virtualni glasovni asistenti imaju nisu vezani za poteškoće u govoru, već za naglaske i dijalekte koji ljudi s raznih krajeva svijeta imaju, također otežavaju asistentu razumijevanje korisnikovog govora. Upravo zbog svega navedenog, budući cilj tvrtki koji razvijaju virtualne glasovne asistente trebala bi biti povećana inkluzivnost prema osobama koje imaju drugačiji govor od neke ustaljene društvene norme na nekom području.

Svakako bi trebalo spomenuti i potencijalne sigurnosne rizike korištenja virtualnih glasovnih asistenata. U radu pod nazivom *Alexa, Can I Trust You?* (2017) Hyunji Chung, Michaela Iorga, Jeffrey Voas i Sangjin Lee navode četiri sigurnosna problema vezana uz virtualne glasovne asistente. Prvi potencijalni problem je nedopušteno pregledavanje interakcije određenog korisnika s virtualnim glasovnim asistentom. Iako virtualni glasovni asistenti koriste enkriptirane mrežne veze, analize su pokazale kako komunikacija između virtualnog glasovnog asistenta i servera s kojim su povezani nije uvijek u potpunosti provedena sigurnosnim protokolima. Pokazalo se da mnogi uređaji ne koriste enkriptirane veze tako otvarajući mogućnost detekcije uređaja s instaliranim virtualnim glasovnim asistentom unutar mreže. Iako je većina komunikacije između asistenta na određenom uređaju i njegovih servera ipak enkriptirana koristeći HTTPS protokol, razne mogućnosti strojnog učenja namijenjene klasifikaciji mrežnog prometa mogu i dalje otkriti brojne informacije o korištenom uređaju. Tako primjerice može biti otkrivena veličina izmijenjenih podatkovnih paketa kao i brzina njihova prijenosa te ostali obrasci prisutni u enkriptiranom internetskom prometu mogu biti iskorišteni za identificiranje statusa uređaja na kojem je virtualni glasovni asistent korišten, kao na primjer, je li uređaj u stanju mirovanja ili u aktivnoj upotrebi, kada je navedeni uređaj uključen ili isključen, razgovara li korisnik trenutno s asistentom ili ne, sluša li glazbu preko asistenta i ostale slične aktivnosti. Drugi sigurnosni rizik je povezanost virtualnih glasovnih asistenata s kućanskim uređajima i ostalim objektima u stambenom prostoru. Ova mogućnost otvara brojne opasnosti vezane uz hakerske napade kao što su, primjerice napadi koji onemogućuju korištenje određenih karakteristika nekog uređaja (eng. *denial-of-service attack*) ili iskorištavanje funkcije *always on* (uvijek uključen) kako bi se omogućilo direktno prisluškivanje razgovora (tj. svih zvukova koji okružuju korišteni uređaj) osoba koje su u neposrednoj blizini uređaja na kojemu je instaliran virtualni glasovni asistent. Također postoji teoretska mogućnost da haker uspije kontrolirati virtualni glasovni asistent putem drugog uređaja kao što je pametni zvučnik ili intercom. Treći sigurnosni rizik su zlonamjerne glasovne naredbe izvršene od strane pojedinaca koji uspiju imitirati glas određenog korisnika virtualnog

glasovnog asistenta. Bez obzira na to što mnogi glasovni asistenti danas nude opciju za učenje i pamćenje glasa pojedinog korisnika, osobe sa sličnom bojom glasa ili intonacijom ponekad mogu uspjeti neovlašteno koristiti tuđeg virtualnog glasovnog asistenta. Ovo omogućava brojne zlonamjerne radnje kao što su, na primjer, otključavanje pametnih vrata bez dopuštenja vlasnika ili kupovina sadržaja na internetu preko asistenta. Sličan, ali mnogo opasniji rizik je i postojanje skrivenih zvučnih poruka čije su zvučne datoteke izmijenjene sa svrhom da budu gotovo nečujne ljudskim ušima, ali uređaji na kojima se nalaze virtualni glasovni asistenti ih i dalje mogu registrirati. Dosad nema konkretne evidencije da je spomenuta metoda ikada korištena, ali je ova mogućnost otkrivena istraživanjem sigurnosnih slabosti virtualnih glasovnih asistenata na sveučilištu Berkeley u Kaliforniji (Smith, 2018). Četvrti sigurnosni rizik je nepostojanje potpune kontrole nad vlastitim glasovnim podacima od strane korisnika virtualnih glasovnih asistenata. Komercijalne tvrtke kao i hakeri mogu ostvariti pristup bazama podataka gdje su spremljeni razgovori svih korisnika određenog virtualnog glasovnog asistenta tako otvarajući mogućnost zloupotrebe informacija koje se nalaze u spomenutim snimkama.

5. Prepoznavanje govora i obrada jezika u virtualnim glasovnim asistentima

U ovom poglavlju bit će ukratko objašnjene osnove razvoja današnjih sustava za prepoznavanje govora i obrade prirodnog jezika. Prva tehnologija za prepoznavanje govora razvijena je 1952. godine u Bell laboratoriju. Bila je sposobna razumjeti samo brojeve od 0 do 9 i to izgovorene isključivo od strane jedne osobe. Iako zvuči skromno za današnje standarde, u pedesetim godinama prošlog stoljeća to je bio prvi veliki korak prema razvoju tehnologije prepoznavanja govora općenito. Sljedeći značajan korak u razvoju tehnologije prepoznavanja govora bio je Harpy sustav razvijen sedamdesetih godina prošlog stoljeća na sveučilištu Carnegie Mellon. Harpy je već bio mnogo napredniji jer je mogao prepoznati preko tisuću riječi kao i različite izgovore tih riječi („Behind the mic“, 2014).

Kasnije, osamdesetih godina počinju se koristiti takozvani *Hidden Markov* statistički modeli za matematičku analizu zvučnih valova koji su donijeli do tada najznačajniji napredak te se zbog toga dominantno koriste i danas. Iako imaju brojne prednosti koje su dovele do danas vidljivog napretka tehnologija za prepoznavanje govora, *Hidden Markov* modeli zahtijevaju ogromne količine podataka za učenje (eng. *training data*) što ponekad može biti problem (Padmanabhan i Johnson Premkumar, 2015). Uobičajen postupak za prepoznavanje govora je uzimanje glasovnog zapisa zvučnih valova koji se zatim dijeli na manje dijelove, a računalo ima zadatak identificirati koji je fonem[2] izgovoren u svakom pojedinom djeliću tog zvučnog zapisa. Dok tehnologija za prepoznavanje govora određuje što je točno izgovoreno, tehnologije za obradu prirodnog jezika pokušavaju odrediti kontekst onoga što je izgovoreno (Pisipati, 2017).

Danas je tehnologija prepoznavanja govora dosegla visoku razinu tako omogućujući korištenje iste velikom broju ljudi jer je dostupna na mnogim uređajima koje danas svi koristimo. Bez obzira na to, još uvijek ne postoji model koji može u potpunosti razumjeti značenje bilo kojeg izgovorenog sadržaja te se kao takav postavlja kao cilj za budućnost razvoja tehnologije prepoznavanja govora („Behind the mic“, 2014).

Naučiti računalo kako sustav prirodnog jezika funkcionira pokazalo se kao izrazito kompleksan zadatak, mnogo teži od osposobljavanja računala za obradu logike, matematike ili ostalih

[2] Fonem je najmanja sljedbena jedinica u jezičnome sustavu koja služi za sporazumijevanje tako da razlikuje značenje, iako je sama bez značenja (definicija preuzeta sa mrežnih stranica Hrvatske enciklopedije).

prirodnih i tehničkih znanosti. S druge strane, impresivna je činjenica kako svako ljudsko biće uglavnom bez poteškoća već u svojem ranom djetinjstvu savlada osnove korištenja prirodnog jezika što je još uvijek velik problem za računala, čak i nakon ogromnog napretka. Ono što čini ovaj zadatak toliko kompleksnim za računala jest semantička (značenjska) raznolikost, korištenje ironije i humora, jaki i raznoliki naglasci, ekspresije lica i govor tijela kao cjelina te na kraju emocije koje su temelj ljudske komunikacije. Sve ovo računala za sada uglavnom ne mogu percipirati, ili mogu, ali u vrlo ograničenom opsegu. Upravo zbog svoje multidimenzionalnosti jezik je izrazito kompleksan zadatak za računala („Behind the mic“, 2014). Važno je napomenuti i problem pozadinske buke koji može značajno narušiti mogućnost cjelokupnog sustava da uspješno razumije što je korisnik izgovorio (Krazit, 2010).

Ljudski mozak kao struktura je izrazito pogodan za učenje jezika, stoga su istraživači tehnologija prirodnih jezika došli do zaključka kako bi virtualna struktura ljudskog mozga mogla unaprijediti razumijevanje prirodnog jezika kod računala. Za tu svrhu stvaraju se umjetne neuralne mreže koje imitiraju strukturu neuralnih mreža u ljudskom mozgu zbog kojih je navedeni toliko napredan i kompleksan. Pokazalo se kako su umjetne neuralne mreže puno efikasnije i uspješnije u razvoju razumijevanja prirodnog jezika od strane računala od prijašnje tehnologije statističke analize jezika. Zato 2009. godine korištenje umjetnih neuralnih mreža postaje rašireno kod mnogih tvrtki koje se bave obradom prirodnog jezika. Prednost neuralnih mreža je njihovo samostalno učenje jezičnih obrazaca kroz vrijeme tako povećavajući uspješnost cjelokupnog sustava što je dovelo do još jednog značajnog napretka unutar tehnologije prepoznavanja govora („Behind the mic“, 2014).

6. Primjeri virtualnih asistenata

U sljedećih nekoliko potpoglavlja bit će opisan osnovni način funkcioniranja svakog od odabranih virtualnih glasovnih asistenata, kao i specifičnosti svakoga od njih. Također će ukratko biti opisana platforma i uređaji na kojima su korišteni. U ovom radu odabrano je pet danas najraširenijih virtualnih glasovnih asistenata, a to su Siri, Google Assistant, Cortana, Alexa i Bixby.

6.1. Siri

Siri je virtualni glasovni asistent razvijen od strane tvrtke Apple te je prvi put predstavljen 2011. godine usporedno s predstavljanjem tada novog Appleovog iPhonea 4S s operacijskim sustavom iOS 5. Siri je danas prisutna na svim Appleovim uređajima, među kojima su iPhone-i, iPad-i, Mac računala, iPod-ovi, Apple satovi, Apple TV-i te HomePod pametni zvučnici. Bez obzira na prisutnost na svim ovim uređajima, iPhone je uređaj na kojemu Siri nudi najviše mogućnosti te je ona zbog toga razvijena prvenstveno za njega. Prilikom predstavljanja iPhonea 4S 2011. godine Siri je smatrana najistaknutijom i najzanimljivijom značajkom novoga pametnog telefona od strane kritičara. Iako ju Apple objavljuje tek 2011. godine, njezin razvoj počinje 2005. godine, od strane tvrtke zvane SRI International Artificial Intelligence Center, a ne Applea. Siri kao aplikacija prvi put je objavljena na Appleovoj trgovini aplikacija AppStore 2010. godine od strane tvrtke koja ju je razvila. Apple tada u toj aplikaciji vidi velik potencijal te odlučuje kupiti tu tvrtku i preuzeti razvoj Siri. Apple također odlučuje Siri učiniti ekskluzivnom na tada novom uređaju iPhone 4S blokirajući njeno korištenje na tada prošlogodišnjem iPhoneu 4 uklanjajući je s AppStorea na kojemu je godinu dana prije bila dostupna („History of Siri“, 2018).

Ovo je bilo protumačeno od strane kritičara i publike kao Appleov pokušaj poticanja tada postojećih korisnika iPhonea na nadogradnju na novi model u svrhu što veće prodaje iPhonea 4S. Danas prepoznatljivi glas Siri posudila je glumica Susan Bennett. Apple 2012. godine Siri čini dostupnom na iPodu i iPadu, kasnije 2015. godine na Apple TV-u te 2016. godine na Mac računalima. 2012. godine mnoge konkurentske tvrtke izbacuju svoje inačice virtualnih glasovnih asistenata koje su u početku bile značajno lošije od Siri, što je Apple tada učinilo vodećim u industriji virtualnih glasovnih asistenata. Kasnije se to mijenja, zahvaljujući kontroverznoj odluci koju je Apple donio vezano za Siri. Apple je odlučio da neće prikupljati podatke vezane za upite koje korisnici upute Siri, već će ih zadržati lokalno na uređaju svakog

pojedinog korisnika. Čak i ako neki upiti moraju biti poslani na Appleove servere, to se obavlja tako da se automatski uklanjaju bilo kakve reference na ime korisnika i bilo koje druge identifikacijske podatke. Ovaj proces na isti način se izvodi i danas te se pretpostavlja kako je upravo to razlog gubitka Appleova dominacije na tržištu virtualnih glasovnih asistenata. Appleovi konkurenti su zauzeli poziciju preuzimanja svih korisničkih podataka vezano za interakciju s virtualnim glasovnim asistentom u svrhu razvoja tehnologije asistenta. Ova politika je također izazvala kontroverze, jer pretvara korisnika u proizvod za prodaju. Sve ovo dovelo je stvaranja narativa u tehnološkom svijetu koji tvrdi da je Apple izgubio svoju staru sposobnost inovacije po kojoj je tvrtka bila poznata. Bez obzira na kritike oko funkcionalnosti Siri, Apple nije posustao u svojoj prvotnoj odluci da nikada ne preuzima korisničke podatke vezane za interakciju sa Siri, postavljajući tako privatnost svojih korisnika kao prioritet. Iako je 2011. godine Siri bila pozitivno prihvaćena od strane kritičara industrije pametnih telefona, imala je vrlo skromne mogućnosti („History of Siri“, 2018).

Sljedeća verzija Appleovog operacijskog sustava za sve Apple mobilne uređaje, iOS 6 donosi novosti poput razumijevanja pitanja vezanih za sport. Siri je mogla odgovoriti na upite vezane uz rasporede utakmica, rezultate, informacije o sportašima kao i postignuća pojedinih klubova za određene sportove. Umjetna inteligencija koja pokreće Siri tada također postaje naprednija otvarajući mogućnost odgovaranja na pitanja vezana za usporedbu. Pitanja tipa „*Who's taller, LeBron or Kobe?*“ Siri je tada mogla uspješno obraditi i dati korisniku ispravan odgovor („History of Siri“, 2018).

Novost je bila i mogućnost odgovaranja na pitanja vezana uz filmove što je postignuto spajanjem s bazom podataka popularne stranice Rotten Tomatoes (https://www.rottentomatoes.com/). Pretraživanje restorana po vrsti kuhinje koju poslužuju, lokaciji, cijenama i korisničkim ocjenama također je dodano. Siri je također postala povezana s aplikacijama za Facebook i Twitter tako omogućujući, primjerice objavljivanje statusa na Facebooku isključivo preko Siri, bez prvotnog otvaranja aplikacije. Novost je bila i mogućnost otvaranja bilo koje aplikacije instalirane na korisnikovom iPhoneu. Zbog svega navedenog iOS 6 bio je najznačajnija nadogradnja sustava za Siri jer je donijela više poboljšanja nego bilo koja kasnija verzija (do trenutka pisanja ovog rada) („History of Siri“, 2018).

Predstavljanjem iOS-a 7 2013. godine Apple donosi nova unaprjeđenja za Siri. Najznačajnija promjena bila je u potpunosti izmijenjeno i poboljšano korisničko sučelje na kojemu se prikazuje poruka „*What can I help you with?*“ prilikom aktivacije Siri. Značajna promjena je

odbacivanje Google tražilice i zamjena istog Microsoftovom internetskom tražilicom Bing. Česta kritika Siri bila je njezina nemogućnost odgovaranja na trivijalna pitanja, što Apple pokušava riješiti spajanjem Siri s Wikipedijom za svrhu odgovaranja na takvu vrstu upita. Ovo rješenje nije najbolje implementirano, jer je Siri u velikom broju slučajeva korisniku samo prikazala cjelokupnu stranicu na Wikipediji vezanu za upit, bez davanja kratkog izgovorenog odgovora. Novost je i mogućnost izmjene postavki sustava, kao što je postavljanje uređaja na način rada u zrakoplovu, uključivanje i isključivanje bluetooth-a, podešavanje razine osvijetljenosti zaslona i slično („History of Siri“, 2018).

Sljedeća verzija iOS-a koja izlazi 2014. godine, iOS 8 donosi „*Hey Siri*“ glasovnu naredbu koja može biti izgovorena bilo kada kako bi aktivirala Siri. To znači da iPhone ili neki drugi uređaj mora konstantno imati uključen mikrofon što je povećalo zabrinutost oko rasta trošenja baterije kod nekih korisnika. Ostale novosti u ovoj verziji su integracija aplikacije Shazam u Siri kako bi Siri mogla prepoznavati pjesme te mogućnost kupovine sadržaja na iTunesu direktno preko Siri. Sljedeća, 2015. godina donosi iOS 9 koji Siri obogaćuje kontekstualnom osviještenošću kao i osviještenošću o lokaciji na kojoj se nalazi uređaj. Ove dvije nove karakteristike značajno unaprjeđuju gotovo sve vrste interakcije sa Siri („History of Siri“, 2018).

Sljedeća, deseta verzija iOS-a izlazi 2016. godine i otvara pristup Siri za developere treće strane putem zasebnog API-ja što je omogućilo povezivanje s aplikacijama za izmjenu poruka (primjerice WhatsApp), dijeljenje prijevoza (primjerice Uber ili Lyft) i ostale. Po prvu puta postaje moguće zatražiti Siri da pošalje poruku unutar neke druge aplikacije osim Appleove iMessage aplikacije. Sljedeća godina, 2017. donosi iOS 11 koji Siri obogaćuje poboljšanim muškim i ženskim glasovima koji bi trebali putem ekspresivnijeg izgovora bolje oponašati uobičajeni ljudski govor, a razvijeni su uz pomoć dubokog učenja (eng. *deep learning*). Siri opet dobiva novo korisničko sučelje kao i novu funkciju prevođenja s engleskog jezika na kineski, francuski, njemački, talijanski i španjolski. Velika novost te godine bilo je učenje obrazaca korištenja uređaja od strane korisnika u svrhu boljeg predviđanja što korisnik želi ili misli pod određenim upitom. Ovi podatci su također sinkronizirani putem Appleovog korisničkog računa na sve ostale Apple uređaje koje korisnik posjeduje i koristi stvarajući tako kontinuirano iskustvo („History of Siri“, 2018).

Naredna, 2018. godina donosi iOS 12 i novu aplikaciju zvanu Shortcuts koja omogućuje korisnicima stvaranje sljedova naredbi koje se mogu spremiti pod jednim imenom i kasnije mogu biti aktivirani umjesto izgovaranja nekoliko naredbi uzastopno. Siri postaje sposobna

odgovarati na pitanja vezana uz hranu (primjerice količinu kalorija, vitamine, minerale kao i upit o tome koliko je određena namirnica zdrava), a također je poboljšano i reagiranje na pitanja vezana uz slavne osobe („History of Siri“, 2018).

Siri je dostupna za korištenje na preko 20 jezika. U vrijeme pisanja ovog rada, Siri funkcionira na engleskom, francuskom, arapskom, kineskom (čak nekoliko varijacija kineskog jezika), danskom, nizozemskom, finskom, njemačkom, hebrejskom, talijanskom, japanskom, korejskom, malajskom, norveškom, portugalskom, ruskom, španjolskom, švedskom, thaiju i turskom („iOS Feature Availability“, n.d). Siri najbolje funkcionira povezana s ostalim Appleovim servisima kao što su Apple TV i Apple Music. Iako podržava pokretanje glazbe i na konkurentskim servisima kao što je Spotify, takva funkcionalnost namjerno je ograničena jer Apple postavlja svoje vlastite servise kao prioritet pri korištenju njihovih uređaja.

6.2. Google Assistant

Google Assistant je Googleov virtualni glasovni asistent službeno predstavljen 2016. godine. Tada predstavljen Google Assistant nije u potpunosti novi proizvod, već nadograđena i preimenovana verzija Googleovog prijašnjeg Google Now glasovnog asistenta. U usporedbi s Google Now, Google Assistant ima više korisnički orijentirano sučelje uz znatno poboljšane konverzacijske mogućnosti postignute opsežnijom primjenom umjetne inteligencije. Kontekstualna svijest je najistaknutija značajka Google Assistant-a čineći ga znatno naprednijim od konkurencije (Purewal, 2016). Google Assistant je dostupan na svim Android pametnim telefonima, seriji Google Home pametnih zvučnika, pametnim zaslonima, automobilima, televizorima, Googleovim prijenosnim računalima i tabletima, Android pametnim satovima te iOS uređajima na kojima je dostupan kao aplikacija na AppStore-u. Prvih nekoliko mjeseci nakon predstavljanja Google Assistant bio je ekskluzivno dostupan samo na Google Pixel i Pixel XL pametnim telefonima i Google Home pametnim zvučnicima, dok tek kasnije postaje dostupan na svim ostalim uređajima.

Google tvrdi kako je njihov asistent sposoban odgovoriti na više od milijun pitanja, što uključuje i naredbe. Kao i ostali virtualni glasovni asistenti danas, Google Assistant može obavljati zadatke vezane uz funkcije uređaja na kojemu je instaliran, obavljati naredbe vezane uz komunikaciju, dati informacije o lokaciji na kojoj se korisnik nalazi, odgovoriti na trivijalna

pitanja, kontrolirati glazbenu reprodukciju i *podcast-e*, a dostupne su i brojne značajke namijenjene za zabavu („Google Assistant“, n.d).

Njegova najjača strana je izrazito napredna kontekstualna svijest. Primjerice, izgovaranjem naredbe “*Show me Tom Hanks movies*“ i nakon dobivanja odgovora na istu moguće je postaviti sljedeće pitanje tipa „*How tall is he?*“. Ovdje Google Assistant razumije kontekst, tj. razumije da se riječ *he* odnosi na Toma Hanksa te će dati odgovor o njegovoj visini („Google Assistant“, n.d).

Zanimljiva značajka Google Assistant-a predstavljena 2018. godine je odgovaranje na dolazne pozive od strane samog asistenta. Prilikom prikaza obavijesti o dolaznom pozivu korisnik može odabrati opciju „Screen call“ kako bi pokrenuo značajku. Tada Google Assistant odgovara na poziv umjesto korisnika, izgovarajući poruku “*Hi, the person you're calling is using a screening service from Google, and will get a copy of this conversation. Go ahead and say your name, and why you're calling*”. Zatim se pokreće transkripcija poziva koja korisniku prikazuje čitavu konverzaciju između strane koja zove i asistenta. Svrha ove značajke je olakšano prepoznavanje „robotskih poziva“ kao i ostalih vrsta nepoželjenih poziva. Korisnik također ima opciju narediti Google Assistant-u izgovaranje pojedinih fraza tijekom poziva kao što su „*Tell me more*“, „*Is it urgent*“, „*Call me back*“ i slične. Nakon završavanja poziva postoji opcija za prijavljivanje poziva kao *spam-a* u slučaju da korisnik zaključi da se radi o „robotskom pozivu“ ili nekoj drugoj vrsti nepoželjenog poziva. Ova značajka za sada je dostupna isključivo na Google Pixel 3 i Pixel 3 XL pametnim telefonima te nije poznato planira li Google dodati istu u Android operacijski sustav kako bi bila dostupna i na ostalim modelima pametnih telefona (Chokkattu, 2018).

Još jedna također inovativna značajka Google Assistant-a predstavljena 2018. godine je Google Duplex. Google Duplex je dodatak na Google Assistant koji omogućuje dogovaranje posjeta i rezervacija s lokalnim restoranima, frizerskim salonima i sličnim objektima. Duplex funkcionira tako da korisnik Google Assistant-u kaže kada i gdje želi dogovoriti dolazak, a zatim asistent samostalno u pozadini upućuje poziv odabranom objektu i dogovora dolazak. Zanimljiv dio je taj što asistent to čini gotovo u potpunosti identično kao čovjek. Glas koji koristi asistent za ovu svrhu zvuči potpuno prirodno, ponajprije zbog dodanih zamuckivanja (primjerice „Umm“) koja ga čine gotovo identičnim ljudskom govoru. Zbog svega toga se čini da Google Assistant ovime uspijeva proći Turingov test, iako neslužbeno (Leviathan i Matias,

2018). Ovakav uspjeh Googlea na području umjetne inteligencije i virtualnih glasovnih asistenata postavlja ga kao vodećeg na tržištu.

6.3. Cortana, Alexa i Bixby

Cortana je virtualni glasovni asistent razvijen od strane tvrtke Microsoft te je predstavljena 2014. godine. Dostupna je na računalima na kojima je instaliran Windows 10, iPhone-ima, Android pametnim telefonima, Xbox One konzolama i Harman Kardon pametnim zvučnicima. Prvotno je bila dostupna i na Windows Phone mobilnom operacijskom sustavu, no Microsoft 2017. godine odustaje od njegovog daljnjeg razvoja zbog neprestanog zaostajanja za konkurencijom (iOS-om i Androidom). Na Windows računalima Cortani se može pristupiti pritiskom na uvijek dostupnu ikonu Cortane na traci sa zadacima (eng. *taskbar*) („Cortana“, n.d). Cortana nikada nije dostigla ni približno visoku popularnost kao Siri ili Google Assistant, ponajprije zbog trenutnog nepostojanja mobilne platforme na kojoj je nativno dostupna, a Windows računala obično nisu prvi odabir korisnika za razgovor s virtualnim asistentom. Iako je dostupna za besplatno preuzimanje kao aplikacija na iOS-u i Androidu, niti na njima nije postigla značajnu raširenost zbog dominacije Siri i Google Assistant-a na njima. Najveći problem vezan uz Cortanu jest njezino zaostajanje za konkurencijom u funkcionalnosti. Dodajući još i slabiju tehnologiju prepoznavanja govora i znatno lošiji izgovor od konkurenata, razumljivo je zašto Cortana danas ne uspijeva postići popularnost (Bacchus, 2019).

Alexa je virtualni glasovni asistent razvijen od strane tvrtke Amazon. Predstavljena je krajem 2014. godine te je dostupna na Amazon Echo i Amazon Echo Dot pametnim zvučnicima koji su razvijeni od strane tvrtke Amazon Lab126, te ostalim Amazonovim uređajima. Za razliku od konkurentskih virtualnih glasovnih asistenata, Alexa nije dostupna za korištenje na pametnim telefonima, tabletima i računalima. Postoji službena Alexa aplikacija za iOS i Android, no njena namjena je upravljanje Alexom koja je instalirana na nekom od Amazonovih uređaja. Amazon je također omogućio ostalim tvrtkama korištenje Alexe u njihovim uređajima (uglavnom se radi o kućanskim uređajima). Alexa se pokreće izgovaranjem njezinog imena nakon čega počinje slušanje korisnikovog upita te odgovaranje na upit. Alexa je sposobna odraditi većinu uobičajenih zadataka kao i konkurentski virtualni glasovni asistenti, ali je posebno orijentirana na Amazonove servise. Tako je moguće samo preko Alexe kupovati proizvode na Amazonovoj internetskoj trgovini koristeći samo glas. Slično kao aplikacija

Shortcuts kod Siri, Alexa također može pohraniti sljedove naredbi pokrenute samo jednom posebnom naredbom. Posebna karakteristika Alexe jest mogućnost proširenja njezine funkcionalnosti instaliranjem takozvanih vještina (eng. *skills*). Najpopularnije vještine su razne igre koje se mogu igrati isključivo glasom (Wetzel, 2019). Alexa se po kvaliteti često uspoređuje sa Siri i Google Assistant-om. Dok Siri i Google Assistant dominiraju kada je riječ o pametnim telefonima, Alexa nedvojbeno dominira kada je riječ o pametnim zvučnicima.

Bixby je virtualni glasovni asistent razvijen od strane južnokorejske tvrtke Samsung. Službeno je predstavljen u travnju 2017. godine uz tada novi Samsungov pametni telefon Galaxy S8 i S8+. Bixby u vrijeme svojeg predstavljanja nije bio u potpunosti novi proizvod, već se zapravo radi u nadograđenoj i preimenovanoj verziji Samsungovog prethodnog virtualnog glasovnog asistenta pod nazivom S Voice. Bixby je ekskluzivno dostupan na Samsungovim pametnim telefonima najviše klase od 2017. godine nadalje. Među njih spadaju Galaxy modeli S8, S8+, Note8, S9, S9+, Note9, S10e, S10 i S10+. Svi navedeni uređaji dolaze s posebnom fizičkom tipkom smještenom na lijevoj strani uređaja namijenjenom isključivo pokretanju Bixbyja. Uz navedeno, moguće je Bixby pokrenuti i uobičajenim glasovnim naredbama kao i ostale glasovne asistente (primjerice „*Hi Bixby*“). Samsung tvrdi kako je glasovnim naredbama upućenima Bixbyju moguće napraviti skoro sve akcije koje se mogu obaviti i dodirivanjem zaslona telefona (Karner, 2019). Posebna značajka jest Bixby Vision koja omogućuje korištenje kamere za identificiranje teksta, predmeta ili lokacija. U slučaju detekcije određenog premeta, Bixby će ponuditi mogućnost traženja tog predmeta za kupovinu na raznim internetskim trgovinama. Ova značajka je vrlo loše implementirana, jer u velikom broju slučajeva Bixby ne uspijeva detektirati prikazano. Kritičari često ističu kako je Bixby vjerojatno najslabiji od svih danas popularnih virtualnih glasovnih asistenata prisutnih na tržištu zbog tromosti i slabog razumijevanja korisničkih upita i naredbi (Hall i Tillman, 2019).

7. Zaključak

Razvoj tehnologija kao što su umjetna inteligencija i obrada prirodnog jezika kroz brojna desetljeća danas nam omogućuje korištenje mnogih uređaja na način koji je nekada bio potpuno nezamisliv. Umjetna inteligencija kao pokušaj imitiranja obrazaca ljudskog uma unaprijedila je računala, ali i cjelokupno čovječanstvo na jednu posve novu razinu. Kako su računala postajala moćnija, ljudi su imali želju stvoriti nove načine korištenja istih. Jedan od tih načina je upravljanje računalom glasom. Nakon otkrivanja potencijala koji se skrivao iza ovog načina korištenja računala, kreće dugi i izrazito složeni proces učenja računala razumijevanju ljudskog govora, kao i vještine računala da govori. Sve to donosi početkom 21. stoljeća *chatbotove* koji po prvi put daju ljudima mogućnost konverzacije s računalima. *Chatbotovi* su ipak bili izrazito skromnih mogućnosti zbog čega se razvoj nastavlja i donosi virtualne glasovne asistente koji danas konačno donose posve novi način korištenja mnogih uređaja. Zahvaljujući velikom trudu uloženom od strane istraživača umjetne inteligencije, danas svatko od nas u svom džepu ima uređaj s mogućnošću primjene umjetne inteligencije, što nam otvara brojne nove mogućnosti. Umjesto uključivanja uređaja i ulaska u određene aplikacije i korištenje istih za svrhu pronalaska određene informacije, moguće je uz virtualne glasovne asistente isto to obaviti u samo nekoliko sekundi koristeći samo glas. Osobe s raznim vrstama tjelesnih oštećenja mogu si tako olakšati korištenje tehnologije primjenjujući ih. Kao i svaka nova tehnologija, virtualni glasovni asistenti imaju neke nedostatke, poput sigurnosnih rizika i slabe inkluzivnosti prema nekim skupinama. Bez obzira na nedostatke, danas je nemoguće zamisliti pametni telefon bez funkcionalnosti virtualnog glasovnog asistenta jer su isti postali sastavni dio našeg svakodnevnog korištenja tehnologije.

8. Literatura

1. Al-Rifaie, M. M. (n.d.). Loebner Prize. Preuzeto 5.6.2019, s http://www.aisb.org.uk/events/loebner-prize
2. [Apple Explained]. (13.5.2018). *History of Siri* [Video datoteka]. Preuzeto s https://www.youtube.com/watch?v=4ryQTkDWmBg
3. Bacchus, A. (9.4.2019). Cortana's Halo Has Grown Dim. Here's What Happened. Preuzeto 14.6.2019, s https://www.virtualtrends.com/computing/cortana-is-dead/
4. Bradeško, L., Mladenić, D. (2012). „A Survey of Chatbot Systems through a Loebner Prize Competition". Artificial Intelligence Laboratory, Institut Jozef Stefan. Ljubljana, Slovenia
5. Chan, S. (15.1.2019). How Speech Recognition Makes Living with Disability Easier (with Tutorials to Get You Started). Preuzeto 6.6.2019, s https://www.achronicvoice.com/2018/03/18/speech-recognition-disability/
6. Chatbots In 2019: (Almost) Everything You Need To Know. (n.d.). Preuzeto 4.6. 2019, s https://www.drift.com/chatbots/
7. Chokkattu, J. (4.12.2018). Here's how Google's Call Screening A.I. works, and how to use it. Preuzeto 10.6.2019, s https://www.virtualtrends.com/mobile/google-call-screening-how-to-use/
8. Childs, M. (31.10.2011). John McCarthy: Computer scientist known as the father of AI. Preuzeto 4.6.2019, s https://www.independent.co.uk/news/obituaries/john-mccarthy-computer-scientist-known-as-the-father-of-ai-6255307.html
9. Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). „Alexa, Can I Trust You?". *Computer, 50*(9), 100-104. doi:10.1109/mc.2017.3571053
10. Corcoran, M. (16.10.2018). People With Speech Disabilities Are Being Left Out of the Voice-Assistant Revolution. Preuzeto 6.6.2019, s https://slate.com/technology/2018/10/voice-assistants-alexa-siri-speech-disabilities-recognition.html
11. Cortana. (n.d.). Preuzeto 14. lipnja 2019, s https://www.microsoft.com/en-us/cortana
12. Fonem. (n.d.). Preuzeto 11.6.2019, s http://www.enciklopedija.hr/natuknica.aspx?id=20057
13. [Google]. (17.10.2014). *Behind the Mic: The Science of Talking with Computers* [Video datoteka]. Preuzeto s https://www.youtube.com/watch?v=yxxRAHVtafI

14. Google Assistant, your own personal Google. (n.d.). Preuzeto 14.6.2019, s https://assistant.google.com/
15. Hall, C. & Tillman, M. (20.2.2019). What is Bixby? Samsung's smart assistant explained. Preuzeto 14.6.2019, s https://www.pocket-lint.com/phones/news/samsung/140128-what-is-bixby-samsungs-assistant-explained-and-how-to-use-it
16. IOS - Feature Availability. (n.d.). Preuzeto 10.6.2019, s https://www.apple.com/ios/feature-availability/#siri
17. Karner, J. (22.2.2019). Samsung updates Bixby with Routines, Voice Unlock, and other features. Preuzeto 11.6.2019, s https://www.androidcentral.com/bixby
18. Knight, W. (3.2.2016). Legacy of Marvin Minsky Carries On in Artificial Intelligence. Preuzeto 4.6.2019, s https://www.technologyreview.com/s/546116/what-marvin-minsky-still-means-for-ai/
19. Krazit, T. (31.8.2010). Google finding its voice. Preuzeto 11.6.2019, s https://www.cnet.com/news/google-finding-its-voice/
20. Leviathan, Y., & Matias, Y. (8.5.2018). Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. Preuzeto 12.6.2019, s https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html
21. Medeiros, J. (19.4.2018). Voice Assistants are Changing How Users with Disabilities Get Things Done. Preuzeto 7.6.2019, s https://www.modev.com/blog/voice-assistants-are-changing-how-users-with-disabilities-get-things-done
22. Neapolitan, R. E., & Jiang, X. (2018). *Artificial intelligence: With an introduction to machine learning*. Boca Raton, FL: CRC Press, Taylor & Francis Group.
23. Padmanabhan, J. & Johnson Premkumar, M. J. (2015). „Machine Learning in Automatic Speech Recognition: A Survey“. *IETE Technical Review*, *32*(4), 240-251. doi: 10.1080/02564602.2015.1010611
24. Pisipati, U. (30.1.2017). Automatic Speech Recognition, Virtual Assistants and the rise of conversational era. Preuzeto 10.6.2019, s https://medium.com/@udaynag/automatic-speech-recognition-virtual-assistants-and-the-rise-of-conversational-era-1493a224d4eb
25. Purewal, S. J. (4.10.2016). The difference between Google Now and Google Assistant. Preuzeto 10.6.2019, from https://www.cnet.com/how-to/the-difference-between-google-now-and-google-assistant/

26. Ramesh R. (13.8.2017). *What is Artificial Intelligence? In 5 minutes* [Video datoteka]. Preuzeto s https://www.youtube.com/watch?v=2ePf9rue1Ao
27. Rouse, M. (listopad 2017). What is virtual assistant (AI assistant)? - Definition from WhatIs.com. Preuzeto 7.6.2019, s https://searchcustomerexperience.techtarget.com/definition/virtual-assistant-AI-assistant
28. Russell, S. J. & Norvig, P. (1995*). Artificial intelligence: A modern approach*. Upper Saddle River: Prentice Hall.
29. Smith, C. S. (10.5.2018). Alexa and Siri Can Hear This Hidden Command. You Can't. Preuzeto 10.6.2019, s https://www.nytimes.com/2018/05/10/technology/alexa-siri-hidden-command-audio-attacks.html
30. Turing, A. (1950). „Computing Machinery and Intelligence“. *Mind, 49*, 433-460.
31. Wetzel, K. (9.4.2019). What exactly is Alexa? Where does she come from? And how does she work? Preuzeto 10.6.2019, s https://www.virtualtrends.com/home/what-is-amazons-alexa-and-what-can-it-do/
32. What is a Chatbot? - Definition from Techopedia. (n.d.). Preuzeto 5.6.2019, s https://www.techopedia.com/definition/16366/chatterbot

Primjena umjetne inteligencije putem virtualnih glasovnih asistenata

Sažetak

U ovome radu cilj je objasniti razvoj i načine upotrebe virtualnih glasovnih asistenata koji se ponekad također nazivaju i inteligentnim osobnim asistentima. Virtualni asistenti su pokretani umjetnom inteligencijom, tj. inteligentnim ponašanjem demonstriranim od strane strojeva, umjesto ljudi i životinja, kao što je to slučaj s prirodnom inteligencijom. Virtualne asistente možemo definirati kao tip softvera koji vrši funkciju prepoznavanja prirodnog jezika (putem glasovnog zapisa ili tekstualnog unosa) i na taj način vrši naredbe zadane od strane korisnika kao što su, na primjer, odgovaranje na pitanja, izvršavanje jednostavnih funkcija na pametnom telefonu ili upravljanje kućanskim aparatima i ostalim uređajima. Virtualni asistenti su danas dostupni ponajviše na pametnim telefonima, ali sve češće se dodaju i televizorima, pametnim zvučnicima, pametnim satovima te osobnim računalima. Nadalje, ovim radom objasnit će se na koji način pojedina tvrtka realizira vlastitog virtualnog asistenta. Tako će biti riječi o Appleovoj Siri, Googleovom Google Assistantu, Microsoftovoj Cortani, Amazonovoj Alexi te Samsungovom Bixbyju.

Ključne riječi: virtualni glasovni asistent, umjetna inteligencija, pametni telefon, prirodni jezik

Application of Artificial Intelligence through Virtual Voice Assistants

Summary

In this work the main point is to explain the development and ways of using virtual voice assistants which are sometimes also referred to as intelligent personal assistants. Virtual assistants are powered by artificial intelligence, i.e. intelligent behaviour showcased by machines instead of humans and animals which is the case for natural intelligence. Virtual assistants can be defined as a type of software which does the function of recognising natural language (either through voice record or text input) and in that way it performs commands given by the user, such as, answering questions, performing simple functions on a smartphone or managing appliances and other home devices. Virtual assistants are available today mostly on smartphones, but they are becoming more and more common on TVs, smart speakers, smart watches as well as personal computers. Furthermore, this work will show how each company makes and manages its own virtual assistant. So there will be a description of Apple's Siri, Google's Google Assistant, Microsoft's Cortana, Amazon's Alexa and Samsung's Bixby.

Keywords: virtual voice assistant, artificial intelligence, smartphone, natural language