

# Konekcionalizam i duboko učenje o aluziji uma: predviđanje nasuprot zaključivanja.

---

Crvelin, Lovre

Undergraduate thesis / Završni rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Croatian Studies / Sveučilište u Zagrebu, Fakultet hrvatskih studija**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:111:550583>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-24**



Repository / Repozitorij:

[Repository of University of Zagreb, Centre for Croatian Studies](#)





SVEUČILIŠTE U ZAGREBU  
FAKULTET HRVATSKIH STUDIJA

Lovre Crvelin

**Konekcionalizam i duboko učenje u filozofiji uma:  
predviđanje nasuprot zaključivanju**

ZAVRŠNI RAD

Zagreb, 2020.



SVEUČILIŠTE U ZAGREBU  
FAKULTET HRVATSKIH STUDIJA  
ODSJEK ZA FILOZOFIJU

Lovre Crvelin

**Konekcionizam i duboko učenje u filozofiji uma:  
predviđanje nasuprot zaključivanju**

ZAVRŠNI RAD

Mentor: doc. dr. sc. Sandro Skansi

Zagreb, 2020.

## Sažetak

U ovom završnom radu ukratko je prikazana povijest razvoja umjetne inteligencije i pristupa u filozofiji uma i u računarstvu pod nazivom Konekcionizam. Također, obrađen je sukob između dvije velike tradicije u filozofiji uma i povijesti umjetne inteligencije, odnosno sukob između klasičnih logičkih simbolističkih modela tumačenja ljudske kognicije i konekcionističkih modela umjetnih neuralnih mreža. Konekcionizam se u novije vrijeme zove duboko učenje (eng. Deep learning) i koristi se kao jedna od glavnih metoda strojnog učenja sa velikim primjenama u razvitku novih tehnologija.

**Ključne riječi:** konekcionizam, duboko učenje, umjetna inteligencija, simbolizam, logika, filozofija uma

## **Sadržaj**

0. Sadržaj

1. Uvod

2. Povijest umjetne inteligencije i dubokog učenja

3. Konekcionizam

3.1. Zakonitosti i način rada umjetnih neuralnih mreža

3.1.1. Hebbovo pravilo

3.1.2. Delta pravilo

3.2. Konekcionizam odnosno predviđanje nasuprot simbolizmu odnosno zaključivanju

4. Priroda spora

5. Zaključak

# 1. Uvod

Konekcijizam u filozofiji uma, odnosno duboko učenje u računarstvu, jedna je od glavnih metoda strojnog učenja i razvoja umjetne inteligencije danas. Gotovo svakodnevno se susrećemo sa digitalnim pojavama koje možemo okarakterizirati kao sustave umjetne inteligencije, na neke od njih smo se toliko navikli, da čak na njih više ne gledamo kao na sustave umjetne inteligencije. Svakog puta kada idemo nešto pretraživati na internetskim tražilicama, kada idemo označiti fotografiju na društvenim mrežama ili možda dati glasovnu naredbu uređaju, mi koristimo sustave koje možemo okarakterizirati kao sustave umjetne inteligencije. Svakog puta kada konzumiramo neki internetski sadržaj, mi zapravo hranimo računalne arhitekture sa podacima iz kojih sustavi uče i svaki put nam sustavi pokušavaju uštedjeti vrijeme, pružiti bolju uslugu, prikazati relevantnije podatke i olakšati život.

Velik dio sustava umjetne inteligencije koristi se metodom dubokog učenja. Što je duboko učenje? Kako i kada se stvorila tehnologija dubokog učenja? Kakve veze filozofija ima s računalnim inženjerstvom? Koji je cilj ovog rada?

Krenuti ćemo s odgovaranjem zadnjeg pitanja, cilj ovog rada je prikazati razvoj umjetne inteligencije kroz povijest, prikazati dvije velike tradicije u filozofiji uma i u računarstvu, napraviti svojevrsnu zahvalu pionirima razvoja umjetne inteligencije o kojima ni svijet filozofije, ni svijet računarstva ne zna dovoljno.

Da bismo dobro obavili ovu zadaću, moramo odgovoriti na pitanje kakve veze filozofija ima s računalnim inženjerstvom. Da bismo to odgovorili moramo se s računarstva vratiti na filozofiju uma. Prije nego su postojali tranzistori, postojale su discipline poput logike, psihologije, a kasnije i kognitivne znanosti, koje nisu bile neovisne znanosti nego su spadale pod filozofiju. Dakle postojali su filozofi i matematičari koji su pokušavali shvatiti kako funkcionira um i koji su pokušali na idejnoj razini razraditi koncepte ljudske ili samo inteligentne misli, odnosno razmišljanja.

Tu dolazimo do centralnog problema ovoga rada, a to je sukob dvije velike tradicije filozofije uma odnosno umjetne inteligencije – predviđanje nasuprot zaključivanju. Drugim riječima to je sukob konekcijizma i simbolizma. Konekcijizam je drugi naziv za sustave dubokog učenja, dok simbolizam predstavlja klasične logičke sustave u filozofiji uma i računarstvu.

Prvo ćemo prezentirati povijest razvoja umjetne inteligencije, potom ćemo detaljnije obraditi način funkcioniranja konekcionizma odnosno deep learning sustava, a za kraj ćemo razraditi glavni sukob između dvije velike tradicije u filozofiji uma i računarstvu.

## **2. Povijest umjetne inteligencije i dubokog učenja**

Povijest umjetne inteligencije seže mnogo dalje od računarstva, još u 17. stoljeću filozof i matematičar Gottfried Leibniz iznio je dvije ideje u kojima možemo naći “začetak umjetne inteligencije.” (Skansi, 2018) Radi se o idejama koje je on nazvao *characteristica univrsalis* i *calculus ratiocinator*. *Characteristica universalis* bi bio univerzalni unificirani jezik u kojemu bi sva znanost mogla izraziti svoje ideje točno i precizno. Takav jezik riješio bi se probleme i ograničenja prirodnog jezika, a pod pretpostavkom konačnog razvoja takvog jezika, stvorili bi se preduvjeti za *Calculus ratiocinator*. *Calculus ratiocinator* bi bio uređaj koji taj precizan znanstveni jezik koristiti kao metodu za racionalno razmišljanje i zaključivanje.

U tako davnim godinama razmišljati o uređaju koji će moći racionalno “razmišljati” i “zaključivati” bilo je revolucionarno, međutim dok smo zaista došli do uređaja koji simuliraju procese koje možemo okarakterizirati kao inteligentne prošlo je četiri stoljeća.

Definiranje umjetne inteligencije nije jednostavan zadatak. Apstraktnim fenomenima, koji se nastoje objasniti i definirati, svaki istraživač prilazi na određeni način. Jako dobar primjer i razlog zbog kojeg područje umjetne inteligencije nailazi na samom početku na poteškoće je taj što je sam koncept inteligencije apstraktni fenomen problematičan za definiranje. Ipak, okvirnu definiciju umjetne inteligencije ponudili su Russell i Norvig (1996) navodeći da je "Umjetna inteligencija područje istraživanja i stvaranja strojeva sposobnih za takvu vrstu aktivnosti, koja bi, da su je izveli ljudi bila proglašena inteligentnom". Ovakva je definicija ipak preuska i ne obuhvaća čitav spektar

onog što umjetna inteligencija uključuje. Russell i Norvig su stoga sakupili velik broj definicija te uočili kako se one razlikuju po nekim detaljima. Iz tog su razloga zaključili kako se definicije mogu klasificirati u četiri okvirne kategorije. Ove kategorije ujedno su i četiri glavna cilja umjetne inteligencije.

1. Djelovanje poput ljudi: pristup Turingovog testa
2. Mišljenje poput ljudi: pristup kognitivnog modeliranja
3. Racionalno mišljenje: pristup zakona mišljenja
4. Racionalno djelovanje: pristup racionalnih agenata

Kao što smo već naveli, umjetna inteligencija kao disciplina se može opisati kao “područje istraživanja i stvaranja strojeva sposobnih za takvu vrstu aktivnosti, koja bi, da su je izveli ljudi, bila proglašena inteligentnom.” (Russell, Norvig, 1943) Naravno, tu dolazimo to problema ljudske prirode i naše navike da se vrlo lako naviknemo na nove stvari. Kao što razvoj svake discipline ima različite epohe, tako i umjetna inteligencija kroz povijest prolazi kroz različite faze koje gledaju na trenutno razvijenu tehnologiju sa drugačijim stavovima. “Nešto što u datom vremenom smatramo inteligentnim, poput računala koje igra šah, nakon nekog vremena postaje uobičajeno, ništa više od običnog računanja.” (Skansi, 2018)

Umjetna inteligencija i filozofija povezane su od davnih dana, prije nego je umjetna inteligencija kao disciplina bila uopće definirana. Ovdje ćemo dati kratak pregled povijesti razvoja umjetne inteligencija, a zainteresiranog čitatelja upućujemo na (Skansi, 2018) gdje može pronaći više detalja od ovoga što prenosimo.

Filozofija u umjetnoj inteligenciji se bavi stvarima poput znanja, značenja, rezoniranja, reference i sličnih pojava koje još uvijek nismo u potpunosti razumjeli, međutim neki od probabilističkih modela nam danas daju obećavajuće rezultate, unatoč našem nesavršenom razumijevanju kognitivnih i mentalnih procesa. Kasnije ćemo objasniti na koji način neki od konekcionističkih modela funkcioniraju, a sad ćemo se vratiti na povijest i temeljna djela u razvoju umjetne inteligencije.

U 19. stoljeću možemo izdvojiti dva temeljna djela koja su doprinijela razvoju umjetne inteligencije, to su *System of Logic* filozofa Johna Stuarta Milla koja izlazi 1843. godine i u kojoj se prvi puta u povijesti logika istražuje kao manifestacija mentalnih procesa. Taj proces logičkog deduktivnog zaključivanja nazivamo silogizam.



Sljedeća utjecajna knjiga jest George Booleova *Laws of Thought* koja izlazi 1854. godine. Boole u knjizi logiku definira kao skup formalnih pravila. Takvo shvaćanje logike ubrzalo je put te discipline do statusa formalne znanosti. Danas je “logika kao znanost smatrana kao dio filozofije i matematike, s mogućnošću implementacije u računarstvu. Razlika između tih logika nije u tehnikama i metodologiji, već u primjenama” (Skansi, 2018).

Nastavkom na 19. stoljeće i Millov *System of Logic*, početkom 20. stoljeća logika je dalje bila povezana sa ljudskim razmišljanjem, a samim time i inteligencijom. Ako bi pokušali stvoriti uređaj koji će manifestirati umjetnu inteligenciju, logično je bilo da krenemo od logike.

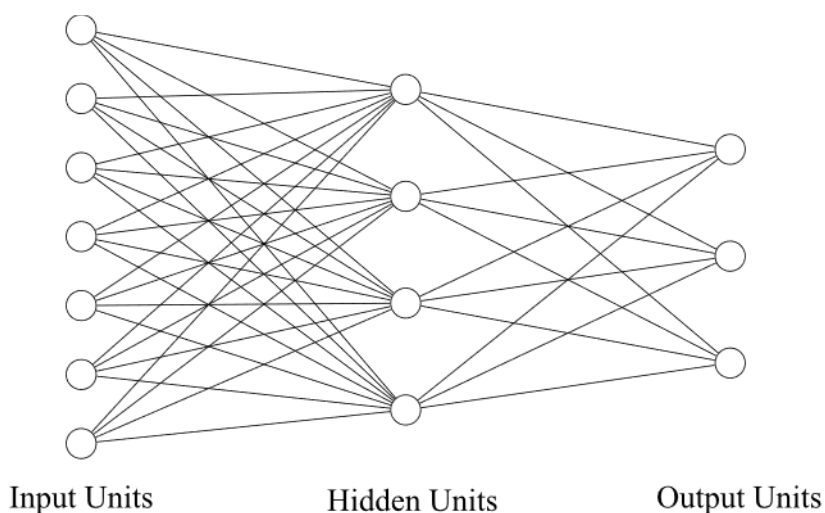
Prve značajne korake razvoja umjetne inteligencije čini “otac računarstva” Alan Turing koji 1936. predstavlja ideju Turingovog stroja. Turingov stroj je apstraktni univerzalni uređaj koji može uzeti upute i obaviti svaki zadatak prema zadanim uputama. Funkcionira na temelju beskonačne trake i seta naredbi koje izvršava ovisno o simbolu na traci. Input u stroj je traka na kojoj postoji simbol, uređaj očitava simbol i s obzirom na definirane naredbe koje se nalaze unutar “kutije” stroja, a sastoje se od odvojenih ćelija, svaka sa svojom naredbom, stroj daje nekakav output. Ovisno o simbolu inputa, stroj odrađuje zadatak i staje, ili mijenja poziciju na traci, briše i stavlja novi simbol na traku. U teoriji, svaki algoritamski proces može se opisati pomoću Turingovog stroja.

1950. godine Turing predstavlja test kojim bi se može odlučiti da li na neko računalo možemo gledati kao na inteligentno ili ne. Test se sastoji od toga da čovjek pet minuta komunicira sa čovjekom i sa računalom. Ukoliko ne može razlučiti što je čovjek, a što računalo, možemo zaključiti da je računalo inteligentno. Turingov test se uz izmjene i prilagodbe koristi i danas.

Sljedeći događaji koji označava veliki milestone u razvoju umjetne inteligencije jest Dartmouth Summer Research Project on Artificial Intelligence. Centralna motivacija konferencije bila opisati sve aspekte učenja i ostalih implikacija inteligencije toliko precizno da se može napraviti uređaj koji će inteligenciju simulirati. Na događaju koji se smatra “rođenjem umjetne inteligencije” (Skansi, 2018), sudjelovali su pioniri istraživanja umjetne inteligencije, John McCarthy, Marvin Minsky, Julian Bigelow, Donald MacKay, Ray Solomonoff, John Holland, Claude Shannon, Nathaniel Rochester, Oliver Selfridgem, Allen Newell i Herbert Simon. Mainstream AI nakon toga postaje simbolistička AI. Simbolistička AI, koju danas također znamo kao GOFAI (good old fashion AI) godinama će sjediti na tronu razvoja umjetne inteligencije, dok je početkom 21. stoljeća ne skine nova tradicija, koju danas znamo kao deep learning odnosno konekcionizam.

### 3. Konekcijonizam

Konekcijonizam je pristup tumačenja uma i kognicije putem umjetnih neuralnih mreža. Neuralne mreže se sastoje od procesorskih jedinica nalik na neurone i veza između različitih slojeva neurona. Neuroni, ili “jedinice”(eng. units), djeluju međusobno na temelju nekoliko neuropsihološki plauzabilnih principa. Njihovo međudjelovanje može biti spojeno na takav način da dobiju kapacitet rješavati kompleksne logičke izračune (McCulloch, Pitts, 1943).



SLIKA 1.

*Prikaz jednostavne neuralne mreže sa skrivenim slojem (Buckner, Cameron and James Garson, "Connectionism", The Stanford Encyclopedia of Philosophy)*

Povijest konekcijonizma, odnosno dubokog učenja, počinje znanstvenim radom Waltera Pittsa i Warrena McCullocha pod nazivom *A Logical Calculus of Ideas Immanent in Nervous Activity* izdan u časopisu *Bulletin of Mathematical Biophysics*.

Priča o Pittsu i McCullochu počinje 1942. kada je McCulloch, tada profesor na Sveučilištu u Chicagu, u svoj dom primio Pittsa, koji je dobio posao na Odjelu Psihijatrije Sveučilišta u Chicagu, a bio je beskućnik.

Pitts je sa 12 godina, skrivajući se u knjižnici nakon što je pobjegao od kuće, pročitao *Principiu Mathematicu* filozofa Bertranda Russela. Nakon toga ga je kontaktirao i Russel ga je pozvao da dođe studirati na Cambridge, što nije bilo moguće s obzirom da je Pitts bio dijete. Međutim, par godina nakon, Russel je imao predavanje u Chicagu i tada su se imali prilike upoznati uživo. Russel je rekao Pittsu da ode posjetiti kolegu logičara Rudolpha Carnapa koji je Pittsu dao knjigu *Logical Syntax of Language* koja je imala velik utjecaj na Pittsa u godinama koje su pratile.

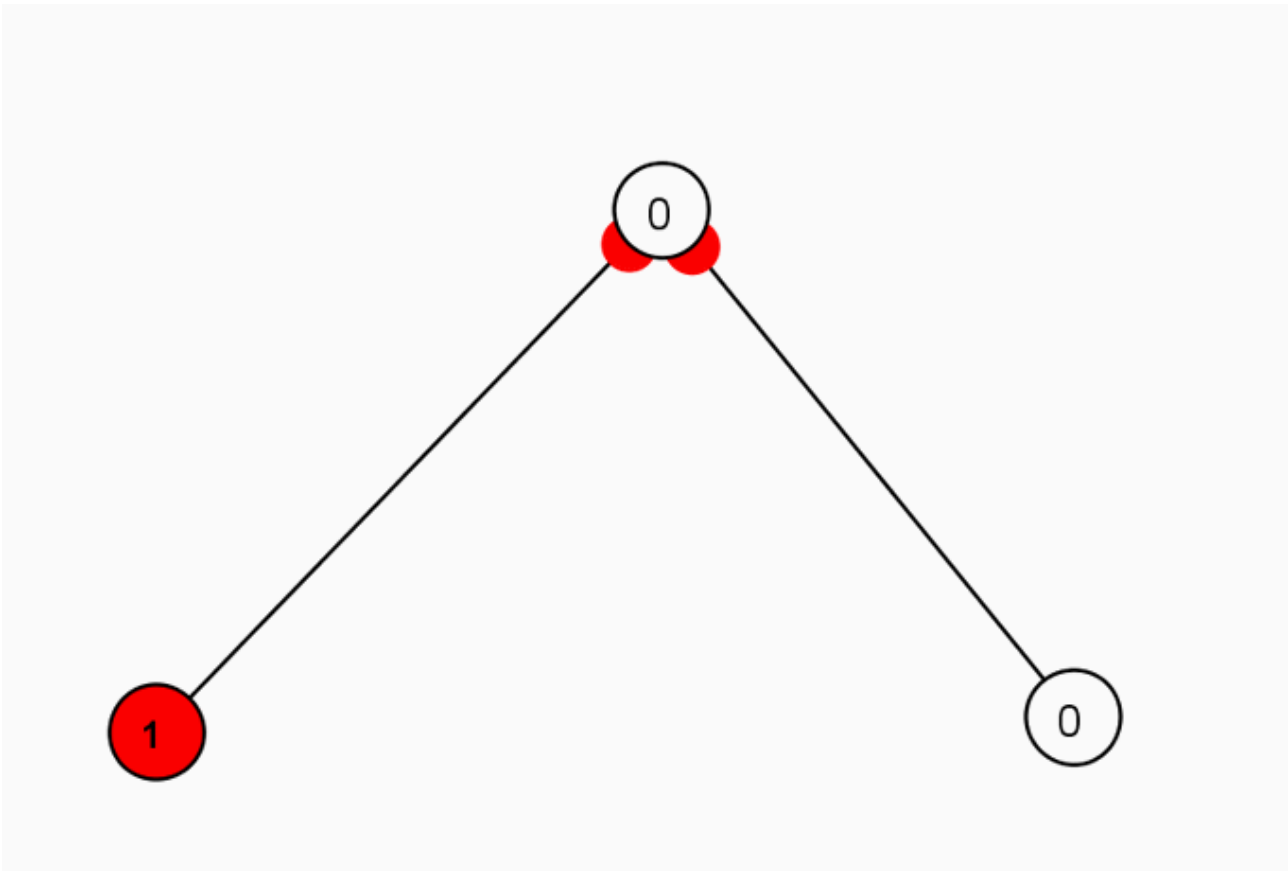
Carnap je iskoristio svoje akademski utjecaj da bi priskrbio Pittsu posao na sveučilištu, kao predavač logike, unatoč tome što Pitts nije imao formalnu diplomu. Pitts se tada mogao posvetiti znanstvenom radu i tada kreće suradnja sa McCullohom. Dijelili su velik interes prema Leibnizu i njegovoj ideji o uređaju koji bi mogao provoditi logičko razmišljanje i dolaziti do zaključaka iz zadanih premisa. Svoje ideje iznijeli su u djelu *A Logical Calculus of Ideas Immanent in Nervous Activity*.

Ubrzo nakon toga Pitts je upoznao matematičara i filozofa, Norberta Wienera, koji je bio stručnjak na području statistike i probabilističkih teorija. Wiener je bio fasciniran neuralnim mrežama jer je znao da kada bi uspjeli modelirati realističnije i kompliciranije mreže, da bi ih mogli staviti u „stroj koji bi tada mogao učiti“ (Geftter, 2015).

### **3.1. Zakonitosti i način rada umjetnih neuralnih mreža**

Pitts i McCulloh su razradu stroja krenuli razvijati na način da su inspiraciju vukli iz ljudske biologije, odnosno na idejnoj razini su pokušali modelirati ponašanje neurona, odnosno neuronskih mreža. Oni su tada napravili distinkciju između input neurona i output neurona. Također, definirali su dva stanja u kojima neuroni mogu biti, pucajućem ili ne pucajućem (eng, firing, non-firing), odnosno oni aktivni koji šalju signal u sljedeći neuron i oni neaktivni koji ne šalju signal. Hoće li neuron pucati ili ne, ovisi o tome da li je signal iz prijašnjih neurona dovoljno jak, odnosno da li je prijeđen određen threshold.

McCulloh i Pitts pokazali su da jednostavnim arhitekturama od 3 neurona, ili “unitsa”, kako su ih tada nazvali, možemo izračunati istinitost operacija konjunkcije, disjunkcije i negacije. Slaganjem više neurona u arhitekturu mogu se rješavati kompleksniji zadaci, a tvrdili su da bi pod pretpostavkom beskonačne trake mogli obavljati bilo koji zadatak koji može obaviti Turingov stroj.



SLIKA 2.

*Konjunkcijska neuralna mreža sa 3 neurona (Jonathan Waskan, Connectionism)*

Svaka procesorska jedinica, odnosno neuron, sadrži aktivacijsku vrijednost( $a$ ), najčešće se to nalazi na skali između 0 i 1. Osim toga, imamo i težinu( $w$ ) koja označava snagu spoja između neurona, ona može biti negativna ili pozitivna.

Utjecaj jednog neurona na drugi neuron može se izračunati i prikazati sljedećom jednačinom:

$$\text{Utjecaj}_{iu} = a_i * w_{iu}$$

Ova jednačina govori da za svaki neuron "i" i za svaki neuron "u", koji je spojen na "i", vrijedi da je utjecaj "i" na "u" jednak umnošku aktivacijskoj vrijednosti za neuron "i" i težini, odnosno snazi

spoja između "i" i "u". Ukoliko neuron "u" ima utjecaj iz više prethodnih neurona, tada se vrijednost utjecaja iz prethodnih neurona na neuron "u" zbraja.

Glavni cilj konekcionističkog istraživanja jest pronalazak pravih težina ( $w$ ) među jedinicama ( $u$ ) da bi neuralne mreže obavile svoj zadatak. Da bi došlo do što boljih rezultata istraživači su razvili razne algoritme po kojima sustavi uče, a generalno možemo razlikovati dvije vrste učenja, nadgledano i nenadgledano učenje.

### 3.1.1. Hebbovo pravilo

Doprinos konekcionističkom pravcu istraživanja predstavio je neuropsiholog Donald Hebb. On je utvrdio da je povezanost između dva neurona jača (odnosno da presinaptički neuron ima snažniji ekscitatorni utjecaj na postsinaptički neuron) onda kada neuroni okidaju zajedno. Ovo pravilo jednostavno je sažeto u njegovim riječima "Neurons that fire together, wire together" (Neuroni koji okidaju zajedno, povezuju se).

Zakovitost koju je Hebb predstavio može se izraziti matematičkom formulom:

Promjena u težini  $w_{iu} = a_i * a_u * \text{irate}$ .

Riječima, težina povezanosti ulazne jedinice "i" prema izlaznoj jedinici "u" treba se promijeniti za iznos jednak umnošku aktivacijske vrijednosti "i", aktivacijske vrijednosti "u" i brzine učenja. Ovo pravilo daje modelima mogućnost da mijenjaju težine na vezama na temelju input-output obrazaca. Hebbov oblik učenja je najpoznatiji oblik nenadgledanog učenja. Kako je svaki input prezentiran mreži, težine između jedinica koje su zajedno aktivne povećavaju se. Suprotno tome, težina veze povezanih jedinica koje nisu aktivirane zajedno smanjuje se.

Ipak, kao pravilo učenja za feed-forward mreže, Hebbovo pravilo nailazi na određena ograničenja. Naročito je neadekvatna činjenica da će učenje jednog input-output para u većini slučajeva interferirati s onime što je mreža već naučila. Ovaj je proces poznat kao katastrofalna interferencija. Drugi je problem taj što je otkrivanje određenih setova težina za izvođenje određenih zadataka izvan dohvata procesa Hebbovog pravila.

### 3.1.2. Delta pravilo

Delta pravilo modificira težine po sljedećoj formuli:

Promjena težine  $w_{iu}$  = brzina učenja \* (željena  $u$  – stvarna  $u$ ) \* stvarna  $i$

Riječima, kako bi se modificirala veza između inputa "i" i outputa "u" delta pravilo računa umnožak razlike između željene i stvarne aktivacije "u", aktivacije "i" i brzine učenja. S obzirom na to, u situaciji kada bi jedinica "u" trebala biti u potpunosti aktivna (a nije) te je input "i" izrazito aktivan, delta pravilo će smanjiti snagu veze "i" prema "u". S druge strane, u situaciji kada "u" treba biti neaktivan, a aktivan je, povezanost "i" prema "u" bit će modificirana u negativnom smjeru.

Frank Rosenblatt nazvao je mrežu temeljenu na delta pravilu Perceptron te je smatrao kako to otkriće predstavlja početak kraja klasičnog procesiranja informacija temeljenog na logici. Perceptron je jednostavna neuralna mreža koja ima dva inputa i jedan output. Ipak, djelo *Perceptrons* Minskya i Paperta (1969) ukazalo je na određena ograničenja u radu perceptrona. Upravo je ovo djelo utjecalo na ukidanje fondova za istraživanje konekcionizma čime su se našli u kompetitivnom zaostatku u odnosu na klasiciste tijekom sljedećeg desetljeća.

Delta pravilo važno je za naglasiti jer predstavlja pojednostavljeni oblik pravila koje čini podlogu za većinu suvremenog konekcionističkog istraživanja. Zatišje u području konekcionizma došlo je kraju otkrićem generaliziranog delta pravila. Ovo pravilo omogućilo je mrežama sa jednim ili više slojeva da izvede one zadatke koje su Minsky i Papert smatrali nemogućim za izvesti. Naime, delta pravilo koristi grešku rezultata (eng. error score), odnosno razliku između stvarne i željene aktivacijske razine izlazne jedinice, te razinu aktivacije dolazne jedinice kako bi se utvrdilo koliko treba promijeniti određenu težinu. Generalizirano delta pravilo funkcionira na isti način za vezu između zadnje skrivene jedinice te izlazne jedinice. U mrežama s više slojeva kada ta veza ide prema drugoj skrivenoj jedinici, pravilo izračunava koliko skrivena jedinica doprinosi ukupnoj pogrešci signala (eng. total error signal), odnosno sumu pogreški signala individualnih izlaznih jedinica, umjesto pogreške signala određene jedinice. Dakle, generalizirano delta pravilo prilagođava vezu jedinice iz ranijeg sloja na skrivenu jedinicu na temelju aktivnosti te prijašnje jedinice te doprinosu kasnije

jedinice u ukupnoj pogreški rezultata. Izuzetna važnost ovog pravila je ta što omogućava backpropagaciju (eng. backpropagation) u kojoj signal greške propagira unazad kroz slojeve mreže kako bi modificirao težine na vezama.

Backpropagation je najčešće korištena metoda nadgledanog (eng. supervised) strojnog učenja. "Nadgledanje" vrše baze podataka s kojima smo nahranili sustav kada smo ga počeli trenirati, a koje sadrže broje primjere već riješenih problema. Na primjer, ako želimo naučiti neuralnu mrežu da razlikuje mačke od psa, u sustav ćemo unijeti velik broj slika sa psima i mačkama na kojima će već biti označeno je li na slici pas ili mačka. Nakon što output neuron izbaci rješenje kod nekog primjera iz trening podataka, mreža će se konzultirati sa već označenim slikama i onda će greške kroz sustav backpropagacije modificirati težine na vezama.

### **3.2. Konecionizam odnosno predviđanje nasuprot simbolizmu odnosno zaključivanju**

Klasicistički pristup ljudskoj kogniciji temelji se na unutarnjim reprezentacijama simbola. Informacije su reprezentirane simbolima, kao što je to slučaj kad nešto zapisujemo u bilježnicu. Klasicisti smatraju da je kognicija analogna digitalnom procesiranju. Kao što u računalima reprezentiramo informacije putem stringova simbola, tako po klasicistima možemo promatrati i ljudski um. Stringovi se stvaraju u određenoj sekvenci prema instrukcijama simbolističkog programa, odnosno sintaksa sadrži pravila o tome kako simbole treba sastavljati.

Prednost simboličkih modela jest u tome da može savladati logičke operacije, dok konekcionistički sustavi neuralnih mreža imaju poteškoća u tom pogledu. Uzmimo za primjer igru poput šaha, simbolistički modeli mogu vrlo precizno i relativno brzo definirati naredbe i opcije, i model će moći "igrati" šah, dok će konekcionističkim modelima takav zadatak biti puno teži. U simbolizmu postoji baza eksplicitnog znanja koje je uređeno na određeni način. Sustav mora imati pravila koja omogućuju pretraživanje baze znanja, izvođenje zaključaka i pohranjivanje znanja. Također, treba imati pravila za dozivanje znanja te uvrštavanje novih pojmova u bazu. Ova vrsta arhitekture predstavlja cjelovitu teoriju ljudske kognicije. Uključuje dugoročna spremišta za deklarativno i proceduralno znanje koje je međusobno povezano. Ta su spremišta povezana i sa vanjskim svijetom preko radnog pamćenja koje je ograničeno. Deklarativno znanje ustrojeno je kao semantička mreža,

a proceduralno u obliku produkcijskih pravila. Radno pamćenje u takvom sustavu predstavlja dio dugoročnog pamćenja koje je trenutno aktivno. U takvom sustavu postoje algoritmi za pretraživanje, pohranjivanje i dozivanje znanja kao i algoritmi za učenje, odnosno stvaranje novog znanja.

Jedna od osnovnih razlika simbolizma u odnosu na konekcionizam jest ta da ne prihvaća pretpostavku da se svo znanje nalazi u vezama među jedinicama odnosno neuronima. Za konekcioniste znanje je implicitno u strukturi modela koji se njime koristi, odnosno obrazac aktivnosti procesirajućih jedinica, dok je prema simbolizmu znanje eksplicitno, odnosno u simbolima. Dakle, konekcionisti smatraju kako informacije nisu spremljene simbolički, već su umjesto simbola važne težine, odnosno snaga povezanosti između jedinica neuralne mreže.

Postoje određene klase problema koje se teško mogu riješiti na klasičan, serijski način. Primjer jest problem optimizacije, a konekcionističke se arhitekture najprikladnije koriste za prepoznavanje govora, slika, modeliranje radnji, navigacije, motorne kontrole i sličnih metoda interakcije sa svijetom.

## **4. Priroda spora**

Priroda spora konekcionizma i simbolizma jest u sljedećem. Treba li se fokusirati na kognitivne kompetencije višeg reda te se fokusirati na rigidna pravila, simboličke reprezentacije i sekvencijalne procese (Fodor, Pylyshyn, 1988) ili da se fokusira na kompleksne detalje inteligentnog ponašanja i da se umjesto simboličkog formalizma odaberu statistički opisi (Smolensky, 1987).

Ako pogledamo ljudski mozak sa neurološke strane možemo uočiti da se on doista sastoji od velikog broja moždanih stanica, neurona, koje su povezane sa sinapsama na druge moždane stanice. Pokušati modelirati ljudsku kogniciju uz inspiraciju iz neurologije čini se vrlo plauzabilno. Postoje sličnosti između konekcionističkih umjetnih neuralnih mreža i mozga, uzmimo za primjer činjenicu da u slučaju propadanja određenih neurona u neuralnim mrežama ne dolazi do prestanka funkcije, već samo do slabljenja točnosti rezultata. Slično, ako osobi uklonite dio mozga, ona neće izgubiti dio sjećanja ili znanja, nego će “znati” iste stvari, ali maglovitije. Također, konekcionističke



neuralne mreže mogu funkcionirati uz input informacije koje su pune šumova, dok će u simbolističkim sustavima mali odmak od predviđenog inputa zapravo uzrokovati pad cijelog sustava.

Filozofija uma stoljećima pokušava doći do razumijevanja o tome kako se koncepti i kategorizacija odvija u našim umovima, Danas je uglavnom prihvaćeno da nužni i dostatni uvjeti za definirati neku stvari brzo dolaze do problema i da se teorijom nužnih i dostatnih uvjeta ne mogu kategorizirati stvari u svijetu, odnosno u ljudskom umu. Uzmimo za primjer životinju poput pande, netko može reći da je to velika životinja sa crno bijelim krznom, bijelom glavom i crnim točkama oko očiju, ali onda se postavlja pitanje što je sa albino pandama? Albino pande definitivno ne ulaze u takav opis, dakle računalo je ne bi moglo prepoznati na takav način, a vrlo je vjerojatno da ni organski mozak tako ne funkcionira. Ono čemu bi kognicija mogla biti sličnija kod ovakvih primjera jest prepoznavanje po sličnosti nekom prototipu ili pripadnosti nekoj kategoriji. Upravo kod ovakvih slučajeva konekcionistički modeli ostvaruju dobre rezultate. Prepoznavanje sličnosti i kategorizacija, prepoznavanje uzoraka u velikim bazama podataka jača je strana neuralnih mreža, dok bi rješenje za takve probleme vrlo teško mogli naći i izraziti putem klasicističkih logičkih pravila i sintakse.

Naravno, konekcionistički modeli imaju problema koje ne smijemo zapostaviti. Neke od osnovnih funkcija mozga poput utjecaja hormona na ponašanje neurona ili klasifikacija neurona po određenim skupinama nisu dio konekcionističkog istraživanja. Također, ako promatramo algoritme učenja poput backpropagationa, nije nam u potpunosti jasno na koji način bi organski mozak obavljao takve funkcije vraćanja u nazad. Međutim, postoje određena istraživanja u kojima su predstavljene višeslojne neuralne mreže sa pojednostavljenim dendridima u kojima postoji error-driven sinapske adaptacije mreže koje djeluju prema željenom rezultatu (Sacramento, Ponte Costa, Bengio, Senn, 2018). Nadalje, kada osoba ugleda neku stvar, dovoljno je da je vidi jednom da nauči što je to, da je sposobna klasificirati tu stvar. S druge strane, umjetnim neuralnim mrežama potrebno je mnogo ponavljanja i treninga da bi naučili što je stvar koja im se prikazuje.

Kritika Fodora i Pylyshyna temelji se na činjenici da ljudsko mišljenje funkcionira na temelju očuvanja istine. Drugim riječima, ako su nečija inicijalna vjerovanja istinita, onda bi vjerovanja koja iz njih proizlaze također trebala biti istinita. Ako mišljenje ne bi bilo „truth-preserving“ na takav način, tada mišljenje ne bi imalo smisla (Fedor, 1987.) Konekcionistički modeli očuvanje istine mogu postići jedino uz korištenje klasičnih simbolističkih arhitektura (Fodor, Pylyshyn, 1988).

Sljedeća stvar koja je srž kritike Fodora i Pylyshyna jest njihova tvrdnja da klasične arhitekture mogu modelirati pojave kao što su produktivnost i sistematizaciju misli. Značenje rečenice se, po njima, nalazi u pojedinim sastavnicama rečenice i u njihovom redoslijedu. Onaj koji nauči sintaksu i semantička pravila prirodnog jezika je sposoban proizvesti neograničen broj gramatički različitih rečenica. Također, onaj koji zna jezik ima sistematsku sposobnost razumjeti i reproducirati rečenice, može proizvesti i razumjeti njihove sistematske varijante. Fodor i Pylyshyn (1988) tvrde da produktivnost i sistematizacija jezika implicira produktivnost i sistematizaciju misli, a da je najbolje objašnjenje rekurzivna i kombinatorna sintaksa i semantika, odnosno simbolistički modeli.

Konekcionizam i simbolizam nisu nužno u proturječnosti, implementalisti primjerice tvrde da je moguće da su za obavljanje logike višega našemu umu potrebne reprezentacije. Oni su svoje istraživanje usmjerili u stvaranje uređaja koji će simboličko procesiranje moći svesti na djelovanje neuralnih mreža.

U raspravi o jeziku, pojavila se rasprava o urođenim sposobnostima. Na pitanje o jeziku simbolisti spočitavaju konekcionistima da ljudski mozak ima predispozicije za učenje jezika te da to ne može biti samo probabilistički sustav. Nedostatak prerceptualnih stimulansa, po nativistima implicira genetičke predispozicije učenja jezika (Chomsky, 1965). Nativisti tvrde da je asocijacija ideja, mehanizam koje su predložili tradicionalni empiristi pre slaba da bi podržala sposobnosti višega reda. Učenje jezika za njih ne može biti po čisto asocijativnom procesu.

Međutim ideja da ljudski mozak evolucijskim putem dođe do nekih predispozicija ne trebaju biti u kontradikciji sa konekcionistima. Štoviše, na evolucijski napredak mozga može se gledati kao na podešavanja težina na modelima da bi znanje koje se treba naučiti postalo lagano, u našem evolucijskom procesu je to jezik.

Uspjeh određenih konekcionističkih modela daje vjetar u leđa empiristima koji smatraju da svo znanje dolazi isključivo iz vanjske percepcije. Ako bi uspjeli u potpunosti razviti modele i demonstrirati dokaze da putem isključivo konekcionističkih modela možemo naučiti jezik, onda bi to imalo zanimljive implikacije na raspravu između racionalista i empirista.

Ekstremni konekcionisti smatraju da klasičan simbolistički pristup nije u sposobnosti riješiti pojave ljudske inteligencije poput shvaćanja konteksta, rezoniranja, asocijacija. Oni žele u potpunosti ukloniti klasičare iz kognitivne znanosti jer smatraju da su usredotočeni na krive stvari.

Konekcionisti tvrde da su reprezentacije zapravo sub simboličke i da analiza sastavnica simbola ne dovodi do željenih spoznaja.

Fokus konekcionista na mišljenje se ne svodi u očitavanju simbola, kao što ne postoji neuron za svaku pojedinačnu stvar, nego se mišljenje manifestira kao rasprostranjena aktivnost po korteksu koja ima neki uzrok. Istraživanja su pokazala da je umjetna neuralna mreža za različite vrste riječi počela stvarati različite uzorke aktivacije na skrivenim slojevima mreže. Upravo u tim uzorcima treba tražiti temelje mišljenja. (Buckner, Cameron, Garson, 2019).

Konekcionisti tvrde da je ključ u uzorcima. Odnosi među reprezentacijama se kriju u sličnostima i razlikama uzoraka. Ovakve distribuirane reprezentacije mogu doprinijeti raspravama o značenju. U simbolističkim reprezentacijskim sustavima značenje se predstavlja u simbolima, strukturi simbola i sintaksi. U distribuiranim reprezentacijama interne sastavnice reprezentacije nose informacije o čemu su (Clark, 1993).

Naravno, ovo pitanje možemo proširiti i na filozofske rasprave o dualizmu, postoji li mentalni odnosno duhovni svijet ili samo materijalni. Ako postoje oba, u kakvom su odnosu, da li je mentalni svijet podskup materijalnog ili materijalni mentalnog? Može li reprezentacija biti u mentalnom svijetu, a fizičko pucanje električnih impulsa iz jednog neurona u drugi zapravo vrsta hardwarea koji omogućava energiji da stvara svijet, odnosno našu svijest? Mogli bi postaviti još mnogo pitanja na koja postoje različiti odgovori i prigovori, a koja bi mogla imati implikaciju na stvaranje uređaja koji će moći razmišljati, međutim to ostavljamo za druge radove.

## **5. Zaključak**

U ovom radu fokusirali smo se na konekcionizam odnosno duboko učenje u filozofiji uma, ukratko smo prikazali povijest umjetne inteligencije, obradili smo dvije velike tradicije u filozofiji uma i računarstvu. Konekcionizam, koji predstavlja korištenje umjetnih neuralnih mreža da bi objasnili i simulirali fenomene ljudske inteligencije, i simbolizam, koji predstavlja klasične logičke simbolističke sustave. Konekcionizam u današnje doba zovu i duboko učenje i jedna je od najčešće korištenih metoda strojnog učenja, sa brojnim primjenama u društvu. Upravo nedavni uspjeh

konekcionističkih arhitektura u brojnim primjenama vratio je interes za istraživanjem neuralnih mreža. Glavni sukob između simbolista i konekcionista nastaje u tumačenju kognitivnih procesa, simbolisti tvrde da se kognicija vrši putem jezika uma, koji barata simbolima po strogim sintaktičkim pravilima, konekcionista svoje modele oblikuju statistički putem umjetnih neuralnih mreža, gdje se do rezultata, odnosno učenja, dolazi tako što se modificiraju snage veza među neuronima i aktivacijska vrijednost svakoga od njih. Korelaciju između mentalnih procesa i aktivnosti neuralne mreže traže u različitim uzorcima aktivacije među slojevima.

Možemo zaključiti da simbolizam i konekcionizam ne isključuju jedan drugog, nego se bave različitim aspektima istog fenomena, ljudske inteligencije. Još uvijek nemamo potpuno razumijevanje o funkcioniranju našeg uma, simbolistički i konekcionistički modeli imaju svoje prednosti i mane, ali niti jedan model još uvijek nije sposoban raditi ono što radi ljudski mozak. Stoga, čeka nas zanimljiva i uzbudljiva budućnost na području razvoja umjetne inteligencije.

## Literatura

Skansi, S. (2018). Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence. Springer.

Russell, S.J., Norvig, P. (2010). Artificial Intelligence: A Modern Approach Third Edition  
Waskan, J. (n.d.). Connectionism. Dohvaćeno iz The Internet Encyclopedia of Philosophy:  
<https://www.iep.utm.edu/>

Buckner, Cameron, Garson, J. (2019) Connectionism. Dohvaćeno iz The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/fall2019/entries/connectionism/>

McCulloch, W., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity  
Bulletin of Mathematical Biophysics

Gefter, A. (2015). The Man Who Tried to Redeem the World with Logic  
Walter Pitts rose from the streets to MIT, but couldn't escape himself. Nautilus, izdanje 021.

Chomsky, N. (1965) Aspects of the Theory of Syntax. Cambridge, MA: MIT

Sacramento, J., Ponte Costa, R., Bengio, Y., Senn, W. (2018) Dendritic cortical microcircuits approximate the backpropagation algorithm. Dohvaćeno iz Advances in Neural Information Processing Systems: <http://papers.nips.cc/paper/8089-dendritic-cortical-microcircuits-approximate-the-backpropagation-algorithm.pdf>

Smolensky, P. (1987) The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn, The Southern Journal of Philosophy.