

# Identifikacija teksta generiranog umjetnom inteligencijom

---

**Biloš, Marina**

**Undergraduate thesis / Završni rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **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:131:210812>

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

*Download date / Datum preuzimanja:* **2024-07-15**



Sveučilište u Zagrebu  
Filozofski fakultet  
University of Zagreb  
Faculty of Humanities  
and Social Sciences

*Repository / Repozitorij:*

[ODRAZ - open repository of the University of Zagreb  
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI  
Ak. god. 2023./ 2024.

Marina Biloš

**Identifikacija teksta generiranog umjetnom  
inteligencijom**

Završni rad

Mentor: dr. sc. Nives Mikelić Preradović, red. prof.

Zagreb, srpanj 2024.

## 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.

*Marina Bieloš*

---

(potpis)



# Sadržaj

Sadržaj.....	ii
1. Uvod.....	1
2. Ljudska ili umjetna inteligencija? .....	2
3. Problem umjetne inteligencije u današnjici .....	3
4. Generiranje teksta.....	5
5. Prednosti i nedostaci različitih UI generatora .....	6
5.1. Jasper A.I.....	6
5.2. Writesonic .....	7
5.3. ClickUp AI .....	9
6. Metode utvrđivanja teksta generiranog od strane UI.....	11
6.1. Ručne metode identifikacije .....	11
6.1.1. Rečenični obrasci .....	11
6.1.2. Identificiranje zastarjelih informacija .....	11
6.2. Softverski alati za otkrivanje teksta generiranog umjetnom inteligencijom .....	12
7. Kako prepoznati UI generiran sadržaj? .....	13
7.1. Tekst koji se ponavlja.....	13
7.2. Neobičan rječnik .....	14
7.3. Predvidljivi uzorci .....	14
7.4. Duljina i struktura rečenice.....	14
7.5. Primjer testiranja .....	15
8. Detektori generiranih sadržaja .....	17
8.1. UI-detektori teksta .....	17
8.2. UI-detektori slike .....	18
8.3. UI-detektori govora .....	19

8.4.	Bihveioralni UI-detektori .....	19
9.	Alati za detektiranje sadržaja generiranog umjetnom inteligencijom .....	21
9.1.	AI Text Detection API.....	21
9.2.	ZeroGPT .....	22
9.3.	DeepAnalyse™ Technology.....	23
9.4.	GPTKit .....	23
9.5.	Writer’s AI content detector.....	24
9.6.	Hive Moderation AI content detector .....	25
9.7.	Content at Scale AI Detector.....	26
9.8.	Raider .....	27
9.9.	Testiranje UI-detektora i njihovi rezultati .....	28
9.9.1.	Testiranje ZeroGPT detektora .....	28
9.9.2.	Testiranje Hive Moderation AI content detektora .....	30
9.9.3.	Testiranje Content at Scale AI detektora .....	31
10.	Rizici korištenja umjetne inteligencije .....	35
11.	Kako označiti sadržaj kojeg je generirala umjetna inteligencija? .....	37
12.	Zaključak .....	39
13.	Literatura .....	40
	Popis oznaka i kratica .....	43
	Popis slika .....	44
	Sažetak .....	47
	Summary.....	48

# 1. Uvod

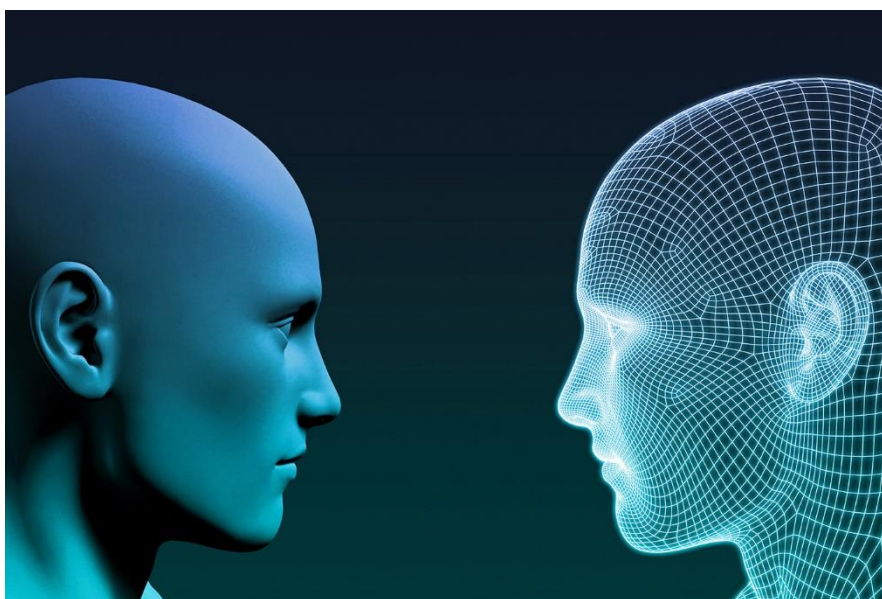
U dvadeset i prvom stoljeću sve se češće spominje pojam umjetne inteligencije koja se također smatra i tehnologijom budućnosti. Postavlja se pitanje kako ona zapravo utječe na čovjekov život. Umjetna inteligencija (*engl. Artificial Intelligence*) predstavlja jedno od tehnološki najznačajnijih dostignuća ovog vremena, mijenjajući način na koji ljudi komuniciraju, obavljaju različite zadatke i razumiju svijet u kojem žive. Ona obuhvaća širok spektar tehnologija i metoda koje računalnim sustavima omogućuju obavljanje zadataka koji prema tradiciji zahtijevaju ljudsku inteligenciju. Taj proces uključuje prepoznavanje govora, vizualnu percepciju, jezično razumijevanje i donošenje različitih odluka. Također podrazumijeva i sposobnost nekog uređaja da oponaša aktivnosti koje inače obavljaju ljudi, u što se ubraja zaključivanje, učenje, planiranje i kreativnost. Područje u kojem je umjetna inteligencija postigla značajan pomak jest generiranje teksta, gdje algoritmi stvaraju koherentne i sadržajno bogatije tekstove koje je teško razlikovati od onih ručno napisanih. Identifikacija teksta (*engl. Text Identification*) kojeg je generirala umjetna inteligencija postaje veliki izazov u digitalnom dobu. S obzirom na sveprisutnije alate za automatsko pisanje, kao što su GPT-3 i GPT-4 modeli koje razvija OpenAI, važno je razviti tehnike i metode za razlikovanje tekstova napisanih ljudskom rukom od onih koje su generirali uređaji. Nedostatak ljudskog nazora može predstavljati izazove, osobito kada se radi o točnosti i ispravnosti podataka. U radu će se analizirati ručne i softverske metode za otkrivanje sadržaja kreiranog od strane umjetne inteligencije, navest će se prednosti i nedostaci različitih UI-generatora i istražiti će se najpoznatiji alati koji se koriste za otkrivanje UI-generiranog sadržaja. Rad će također navesti i mogućnost označivanja sadržaja stvorenog od umjetne inteligencije kako bi se korisnicima pružilo bolje razumijevanje o postojećim tehnologijama povezanim s umjetnom inteligencijom.

## 2. Ljudska ili umjetna inteligencija?

Umjetna inteligencija (UI) i strojevi s vrlo visokim kapacitetom su tradicionalno povezani s ljudima. Ipak, od druge polovice dvadesetog stoljeća i sa ubrzanim porastom i razvojem tehnologije u posljednjih dvadesetak godina, oni su pokazali sposobnost i primjenu učenja na način na koji su to prije mogli samo ljudi (Bennett, 2024).

Vrlo je važno spomenuti da se ljudska i umjetna inteligencija značajno razlikuju i ne mogu u potpunosti zamijeniti jedna drugu. Čak i kada postoje dvojbe oko toga što točno definira umjetnu, a što ljudsku inteligenciju, razlike među njima su jasne.

Tisućama godina unazad vode se rasprave kako najbolje opisati i definirati pojam ljudske inteligencije. Stvoreno je stotine definicija, ali sve one zajedno dovode do krajnjeg zaključka da je ljudska inteligencija biopsihološka sposobnost stjecanja i primjene znanja i vještina. S druge strane, umjetna inteligencija, gdje je poseban naglasak na strojnim algoritmima, nevjerojatno je učinkovita u obradi i integraciji novih informacija i dijeljenju novog znanja među odvojenim modelima umjetne inteligencije. Također, umjetna inteligencija radi brzinom znatno većom od brzine ljudske inteligencije (Bennett, 2024).



*Slika 1.* Ljudska i umjetna inteligencija



### 3. Problem umjetne inteligencije u današnjici

Autori sadržaja se sve više oslanjaju na alate umjetne inteligencije kako bi povećali svoju produktivnost, upravljali radnim procesom i riješili se dugotrajnog i zamornog posla. Međutim, kako strojevi uče obavljati složenije i naprednije zadatke, dolazi do pitanja hoće li zaposlenicima i drugim profesionalcima u području tehnologije ostati dovoljno posla na visoko konkurentom tržištu (Russel, 2022). Velika većina stručnjaka smatra da će do 2035. godine tehnologija automatizirati zadatke koji se ponavljaju kako bi se uštedjelo vrijeme te oslobodilo radnike od jednostavnih zadataka kako bi se usredotočili na zahtjevnije poslove. Iako strojevi mogu obavljati mnoge zadatke brže i preciznije od ljudi, vjeruje se da još uvijek postoji potreba za ljudskom intervencijom, pogotovo u kreativnim profesijama kao što je na primjer, novinarstvo. Radi lakšeg razumijevanja, alati za generiranje teksta mogu provjeriti i ispraviti gramatičke pogreške ili pak dodati određene riječi, ali na čovjeku je da smisli samu temu članka, označi njegovu strukturu i kreira njegov nacrt. Drugim riječima, ljudski je posao generirati ideju, a umjetna inteligencija služi da tu temu materijalizira. Marketinški stručnjaci za svoje sadržaje također mogu koristiti analitičke alate koje pokreće UI kako bi predvidjeli koliko će njihove marketinške kampanje biti uspješne, pratili ključne trendove i mjerili rezultate. Osim toga, umjetna inteligencija ubrzava pisanje oglasa, objava na društvenim mrežama i drugih prijedloga nudeći predloške temeljene na temama i ključnim riječima (Bennett, 2024).

Alati umjetne inteligencije doista su postali toliko moćni da mogu besprijekorno obavljati različite zadatke za koje se prije desetak godina smatralo da ih mogu obaviti samo ljudi. Kao rezultat toga, neka su velika poduzeća smanjila broj zaposlenih. Iako bi neki poslovi mogli zastarjeti, smatra se da će umjetna inteligencija dovesti do novih nekonvencionalnih karijera. Na primjer, većina poslovnih vođa vidi potencijal za uloge kao što su treneri umjetne inteligencije, robotski treneri i napredni podatkovni znanstvenici (Swarup, 2012). Kako se odnosi između ljudi i umjetne inteligencije neprestano razvijaju, profesionalci koji su u bliskom dodiru s tehnologijom imaju veliku prednost na tržištu. Smatra se da će uz pomoć umjetne inteligencije raditi brže i s manje grešaka pa će stoga imati i više vremena posvetiti se kreativnijem dijelu posla.

Ipak, još uvijek postoji mnogo zadataka koje samo ljudi mogu obaviti, a umjetna inteligencija smatra se samo pomagalom, a ne konkurencijom.

## 4. Generiranje teksta

Generiranje teksta (*engl. Text Generation*) proces je kojim umjetna inteligencija stvara pisani sadržaj, istovremeno oponašajući obrasce i stilove ljudskog jezika. Ova aktivnost uključuje stvaranje koherentnog i smislenog teksta koji oponaša prirodnu ljudsku komunikaciju. Generiranje teksta postalo je vrlo važno u različitim područjima, uključujući obradu prirodnog jezika, kreiranje sadržaja, korisničku podršku i pomoć pri kodiranju. Koristi algoritme i jezične modele za obradu ulaznih podataka i stvaranje izlaznog teksta. Sam proces uključuje uvježbavanje UI-modela na velikim skupovima tekstualnih podataka kako bi se naučili obrasci, gramatika i kontekstualne informacije. Ti modeli zatim primjenjuju naučeno znanje za generiranje novog teksta na temelju zadanih upita ili uvjeta (Awan,2023).

U središtu generiranja teksta su jezični modeli poput GPT-a (*Generative Pre-trained Transformer*) i Googleovog PaLM-a, koji su obučeni na ogromnim količinama tekstualnih podataka s Interneta. Ovi modeli koriste tehnike dubokog učenja za razumijevanje strukture rečenica i stvaranje koherentnog i kontekstualno relevantnog teksta, a ključnu ulogu imaju neuronske mreže.

Tijekom procesa generiranja teksta, UI-model prima početni unos poput rečenice ili ključne riječi te koristi svoje naučeno znanje za predviđanje najvjerojatnijih sljedećih riječi ili fraza. Model nastavlja stvarati tekst, održavajući kontekst i koherentnost, sve dok se ne postigne željena duljina ili dok se ne ispune zadovoljavajući uvjeti. Najčešći zadaci koje generatori teksta ispunjavaju su stvaranje novog sadržaja, interakcija s korisnikom u kojoj se na njegov upit nude relevantni odgovori i informacije, generiranje točnog prijevoda na temelju analize velike količine prevedenih tekstova te sažimanje teksta na najvažnije informacije koje su potrebne korisnicima (Awan, 2023).

## 5. Prednosti i nedostaci različitih UI generatora

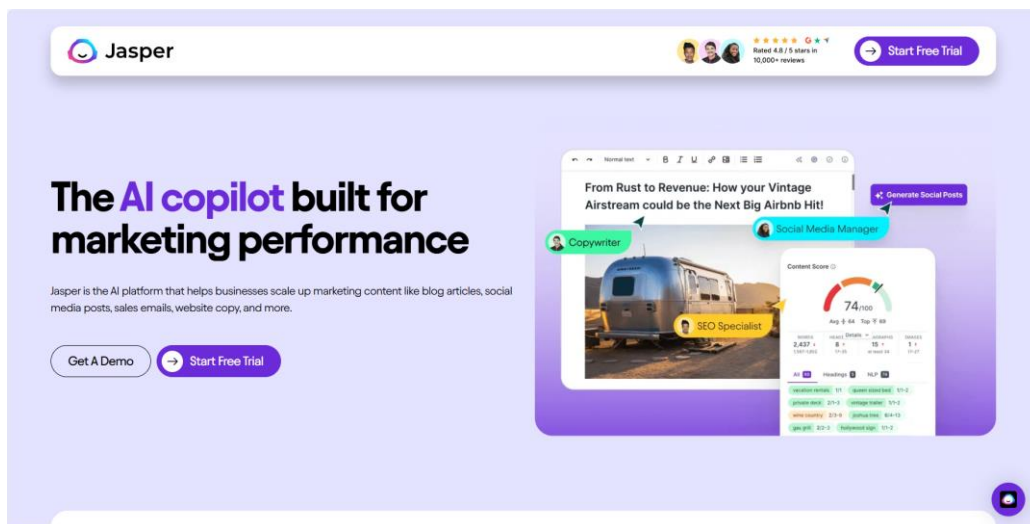
U posljednjih nekoliko godina, generatori umjetne inteligencije postali su ključni alati u različitim industrijama, omogućujući stvaranje sadržaja, analizu podataka i automatizaciju procesa na načine koji su prije bili nezamislivi. U narednim ulomcima, detaljno će se istražiti prednosti i nedostaci različitih UI-generatora, razmatrajući kako njihova specifična svojstva i mogućnosti mogu utjecati na učinkovitost i kreativnost korisnika. Na taj način, cilj je pružiti sveobuhvatan pregled koji će pomoći korisnicima da odaberu najprikladniji alat za svoje potrebe, dok će istovremeno biti svjesni mogućih izazova koje takvi alati mogu donijeti.

### 5.1. Jasper A.I.

Jasper A.I., moćan jezični model, koristi se za generiranje teksta, prevođenje, stvaranje raznovrsnih oblika kreativnog sadržaja i pružanje relevantnih odgovora na postavljena pitanja korisnika. Koristi napredne modele za obradu prirodnog jezika (NLP) za kreiranje visokokvalitetnog teksta. Također, služi kao korisnička podrška te pomaže u širenju marketniga putem e-pošte. Vrlo je dobar u pisanju tekstova i stvaranju različitih vrsta sadržaja, a njihova kvaliteta ovisi o korisničkim preferencijama i zahtjevima (Bretous, 2023). Postoji mnogo razloga zbog kojih je Jasper A.I. dobro poznat. Neki od njih odnose se na lakoću korištenja, dok drugi podrazumijevaju različite formate koje on pruža.

Kao što je već spomenuto, Jasper A.I. ima snažne mogućnosti u kreiranju sadržaja. Može generirati *postove* na blogu, članke, natpise na društvenim mrežama, tekst koji se koristi kod e-pošte i još mnogo toga. Nudi i mogućnosti prevođenja jezika, pisanja marketinških materijala i kreiranja kreativnih opisa za proizvode koji se nude korisnicima. Jednostavan je za korištenje jer sadrži čisto korisničko sučelje s puno opcija. Potrebno je samo kliknuti na ponuđene ikone kako bi korisnici pronašli određeni dokument ili drugu vrstu sadržaja koja im je potrebna. Uz Jasper A.I., moguće je dobiti čitljiv sadržaj. Razlog je taj što program pamti ono što je prije izgenerirao i pokušava izbjeći znakove koji otkrivaju da je sadržaj generiran umjetnom inteligencijom, kao što su fraze koje se ponavljaju i neuredan vokabular (Chaudhary, 2023).

S druge strane, potrebno je spomenuti i nedostatke. Cijene alata su poprilično visoke, kreću se od trideset devet do pedeset devet dolara mjesečno, no alat ne pruža uvijek dovoljno dobre rezultate za izdvojeni novac. Također, odgovori vrlo često mogu biti pogrešni ako mu korisnik ne da dovoljno detaljan upit, što dovodi u pitanje i korisnost alata jer je potreban ljudski doprinos. Jedan od najznačajnijih nedostataka alata je taj što se ne snalazi dobro s tehničkim temama. Na primjer, teško nalazi odgovore vezane uz tehničke specifikacije. Stoga, ako korisnik traži upit o kojem na Internetu još uvijek nema poznatih podataka, Jasper A.I. će se „namučiti” jer ne može pronaći koherentne informacije. Zaključno, Jasper A.I. traži činjenice na Internetu prije pisanja te nema vještine kritičkog razmišljanja. Upravo se iz toga razloga dezinformacije i netočni podaci mogu pojaviti u izgeneriranom tekstu (Chaudhary, 2023).



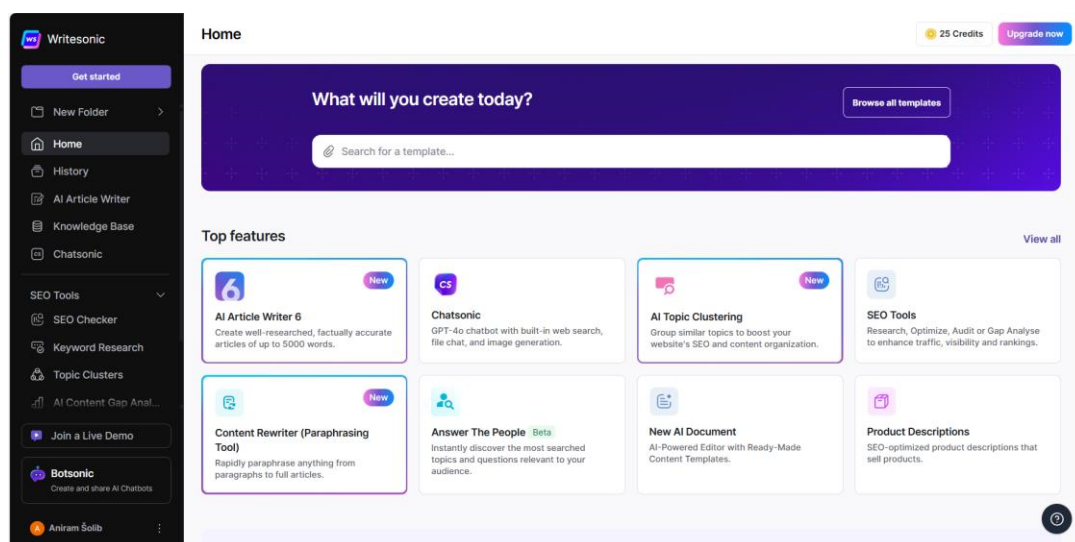
Slika 2. Jasper A.I.

## 5.2. Writesonic

U svijetu stvaranja sadržaja koji se neprestano razvija, Writesonic se ističe kao pokretač inovacija, nudeći paket alata koji se pokreću umjetnom inteligencijom i značajke SEO optimizacije koje autorima i marketinškim stručnjacima omogućuju postizanje izvanrednih rezultata u stvaranju sadržaja. Cilj SEO optimizacije je da web stranica bude bolje rangirana u rezultatima pretraživanja za određene ključne riječi ili fraze koje su relevantne za sadržaj stranice.

Što se tiče prednosti Writesonica, potrebno je spomenuti njegovu učinkovitost. Korisnici, a posebno marketinški stručnjaci, blogeri, pisci i tvrtke, mogu koristiti Writesonic za brzo kreiranje potrebnog sadržaja poput članaka i blog *postova* bez potrebe za angažiranjem velikog broja pisaca ili trošenjem vremena na pisanje. Može generirati tisuće riječi mjesečno, ovisno o planu pretplate. Inuitivan je i jednostavno se koristi, čak i za korisnike koji nemaju mnogo tehničkog znanja. Također, za dobre rezultate dovoljan je minimalan korisnički upit.

Iako Writesonic nudi brojne prednosti, postoje i neke negativne strane koje treba uzeti u obzir. Kvaliteta sadržaja može varirati, ponekad generirajući neprecizne tekstove koji zahtijevaju dodatno uređivanje. Nedostatak originalnosti može biti problematičan za korisnike koji traže jedinstven sadržaj, a generirani tekstovi često trebaju ljudsku provjeru kako bi se osigurala točnost i relevantnost. Writesonic nema dovoljno znanja o tehničkim temama i specijaliziranim područjima pa postoji i rizik od generiranja netočnih informacija. Troškovi naprednih značajki mogu biti visoki do sedamdeset devet dolara mjesečno, a pretjerana uporaba alata može smanjiti produktivnost korisnika. Stoga je važno koristiti Writesonic kao pomoćni alat, a ne kao zamjenu za ljudsku kreativnost i stručnost (Baldikov, 2024).



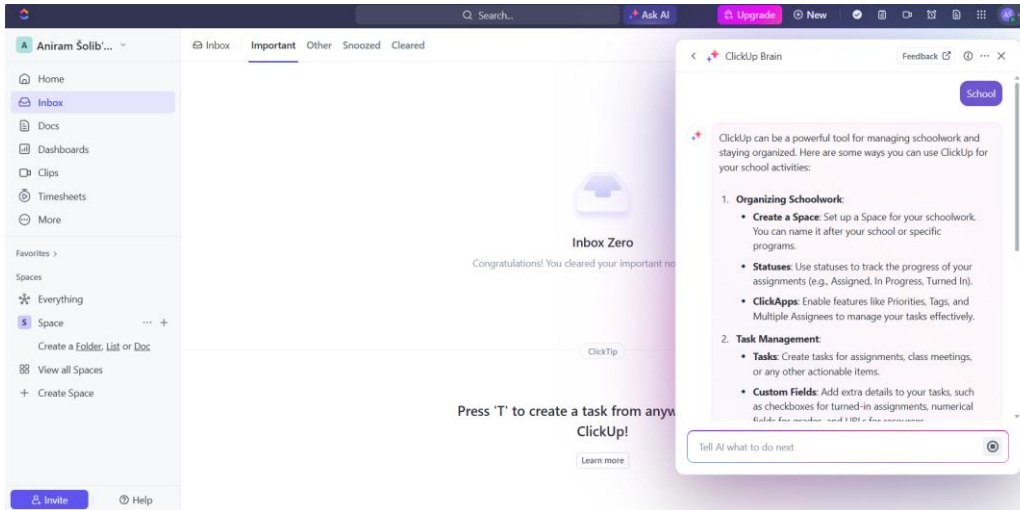
Slika 3. Writesonic

### 5.3. ClickUp AI

ClickUp AI je prilagodljiv i moćan alat za upravljanje projektima koji je namijenjen timovima, ali i drugim korisnicima. Alat omogućuje timovima organizaciju poslova, suradnju i praćenje napretka kroz različite funkcionalnosti te ujedinjuje podatke o projektima tvrtke u jedno *online* rješenje (Manuel, 2024).

ClickUp AI nudi mogućnosti koji obećavaju produktivnost i promjenu načina na koji se umjetna inteligencija koristi u timskoj suradnji te kod upravljanja zadacima. Ističe se svojim funkcionalnostima prilagođenim različitim ulogama, popularan je među timovima različitih veličina, počevši od malih timova do velikih organizacija. Primjeri uključuju tehnološke tvrtke koje upravljaju razvojem softvera i IT projektima, marketinške agencije koje planiraju i prate kampanje, proizvodne tvrtke koje koordiniraju proizvodne procese i upravljaju zalihama, kao i kreativne industrije koje organiziraju dizajnerske projekte i video produkciju. Prednostima se smatraju jednostavan pristup upitima, mogućnosti automatizacije i besprijekorna integracija s tijekovima rada platforme, što omogućuje korisnicima učinkovitije vođenje projekata i timskih aktivnosti. Vrlo dobro sažima tekst te piše i uređuje dokumente i komentare.

Kao nedostatke valja spomenuti činjenicu da generirani sadržaj ostaje na bazičnoj razini. Potrebno je dodatno uređivanje kako bi se tekst učinio osobnijim, što uključuje dodavanje vlastitih riječi i stila kako bi se istaknula autentičnost. Također, testiranjem alata zaključeno je da često daje pogrešne odgovore koji uopće ne postoje na Internetu. To je zabrinjavajuće jer takve greške mogu dovesti do širenja netočnih informacija. Na kraju, potrebno je spomenuti i nedostatke koje nudi besplatna verzija. Korisnici besplatne verzije obično ne dobivaju prioritetnu podršku ili pristup dodatnim resursima poput obuka ili tutorijala. Također, generirani sadržaj u besplatnoj verziji može biti označen vodenim žigom ili ograničen u pogledu mogućnosti formatiranja i izvoza, odnosno korisnici možda neće imati mogućnost izvoza generiranog sadržaja u različitim formatima ili prilagoditi izlazni format prema svojim potrebama (Manuel, 2024).



Slika 4. ClickUp AI



## 6. Metode utvrđivanja teksta generiranog od strane UI

Postoji više načina da se utvrdi je li tekst generiran umjetnom inteligencijom. Oni se mogu kategorizirati u dva glavna pristupa: ručna identifikacija i identifikacija temeljena na softveru.

### 6.1. Ručne metode identifikacije

Istraživanja su pokazala da, iako ne postoji besprijekorna metoda za prepoznavanje teksta generiranog umjetnom inteligencijom, specifične karakteristike često se pojavljuju u takvom sadržaju (Amalia, 2023).

#### 6.1.1. Rečenični obrasci

Tekst generiran umjetnom inteligencijom često ispisuje jedinstvene interpunkcijske uzorke. To se može objasniti kao pretjerana upotreba uobičajenih riječi kao što su „ono”, „to” ili „jest” umjesto specifičnije terminologije. Ove posebnosti proizlaze iz UI-jevih algoritama za predviđanje teksta. Nasuprot tome, tekst koji su u potpunosti oblikovali ljudi obično prikazuje širi raspon fraza i rečenica, uključujući različite stilove, pa čak i sleng (*engl. slang*). Kao rezultat toga, mogu se pojaviti pogreške pri upisu. Još jedna karakteristika koju treba uzeti u obzir je ponavljanje riječi i fraza. Iako su noviji alati umjetne inteligencije dizajnirani da bude kreativniji, sadržaj generiran ranijim modelima često iznova koristi iste ključne riječi i teme s malo raznolikosti (Amalia, 2023).

#### 6.1.2. Identificiranje zastarjelih informacija

Osim analize rečeničnih obrazaca, ispitivanje vremenskog okvira informacija može biti zanimljivo tijekom promatranja izgeneriranog sadržaja. Na primjer, od 2023. baza podataka ChatGPT-a uključuje podatke do siječnja 2022. Paralelno tome, alat generira sadržaj na temelju informacija dostupnih do tog trenutka. Ako članak ovisi isključivo o informacijama koje su nastale prije tog datuma, može se naslutiti da je sadržaj možda generiran umjetnom inteligencijom. Budući da nije moguće pristupiti podacima nakon siječnja 2022., sadržaju generiranom umjetnom inteligencijom možda nedostaju detaljnije ili novo otkrivene informacije. Identificiranje ovih značajki

može pomoći u određivanju je li neki članak u potpunosti ili djelomično stvoren od strane umjetne inteligencije. Razni istraživači i programeri također su stvorili alate za otkrivanje teksta generiranog umjetnom inteligencijom, uglavnom na temelju ovih smjernica (Amalia, 2023).

## **6.2. Softverski alati za otkrivanje teksta generiranog umjetnom inteligencijom**

Postavlja se pitanje postoji li konačan način za prepoznavanje teksta generiranog umjetnom inteligencijom pomoću softverskih alata. Odgovor je delikatan jer trenutno niti jedan alat ne može nedvosmisleno razlikovati tekst koji su proizveli ljudi od onog koji je generirala umjetna inteligencija. Ipak, obrasci o kojima se danas sve više raspravlja mogu pomoći u bržoj identifikaciji.

Identifikacija teksta generiranog umjetnom inteligencijom koristi niz softverskih metoda kako bi se osigurala preciznost i pouzdanost u prepoznavanju takvog sadržaja. Ove metode uključuju analizu ponavljanja, koja detektira učestala ponavljanja ili neobične obrasce u tekstu, kao i lingvističku analizu koja koristi alate za obradu prirodnog jezika kako bi identificirala karakteristike tipične za UI-generirane tekstove. Također se koristi detekcija specifičnih markera koja pronalazi riječi, fraze ili sintaktičke konstrukcije koje su česte u UI-generiranim tekstovima. Dodatno, softver može koristiti specijalizirane alate i API-je za identifikaciju UI-generiranog sadržaja te se može trenirati pomoću strojnog učenja na velikim skupovima podataka kako bi naučio razlikovati ljudski i UI-generiran tekst. Kombinacija ovih metoda omogućuje pouzdanu identifikaciju UI-generiranog sadržaja u različitim kontekstima i industrijskim sektorima (Amalia, 2023).

## 7. Kako prepoznati UI generiran sadržaj?

Utjecaj umjetne inteligencije u različitim područjima je u današnjici široko prisutan. UI-tehnologije sve više oblikuju industrije kao što su zdravstvo, financije, obrazovanje i transport te donose inovacije koje optimiziraju procese, poboljšavaju usluge i pomažu u rješavanju složenih izazova modernog društva. Unatoč svakodnevnim raspravama o njezinim etičkim implikacijama, uloga umjetne inteligencije u stvaranju sadržaja je svakako značajna. Međutim, sadržaj generiran umjetnom inteligencijom može predstavljati i različite izazove (Ward, 2024). Postoje dvije karakteristike unutar teksta koje osiguravaju prepoznavanje sadržaja generiranog umjetnom inteligencijom, a to su zbrkanost (*engl. confusion*) te praskovitost (*engl. burstiness*).

U UI-detekciji, zbrkanost podrazumijeva nerazumljivost ili nesklad u generiranom sadržaju. Tekstovi generirani umjetnom inteligencijom mogu sadržavati nejasne ili nelogične informacije zbog pogrešnog tumačenja konteksta ili problema s obradom prirodnog jezika. Prepoznavanje takve zbrkanosti može pomoći u identificiranju teksta koji nije napisan od strane čovjeka. S druge strane, praskovitost se odnosi na nepravilnu distribuciju sadržaja ili pojmova unutar teksta. Ljudi obično pišu s određenom razinom varijacije i prirodne neujednačenosti, dok UI-generiran tekst može imati neuobičajeno visoke ili niske razine varijacije u određenim dijelovima. Analiza praskovitosti može biti ključna za otkrivanje umjetne inteligencije jer pomaže u razlikovanju prirodnih obrazaca ljudskog pisanja od algoritamskih („Kako točno rade AI detektori?”, 2023). Za nekoliko različitih vrsta sadržaja smatra se da su generirani umjetnom inteligencijom.

### 7.1. Tekst koji se ponavlja

Kada umjetna inteligencija generira tekst, veća je vjerojatnost da će doći do ponavljanja određenih riječi. Bilo da se radi o slučajnom dupliciranju riječi ili izraza, stvara se obrazac koji UI-detektor može otkriti. U stvarnosti, tekst koji su napisali ljudi ima manje ponavljanja te oni koriste raznovrsniji jezik u svakodnevnom govoru (McLean, 2024).

## **7.2. Neobičan rječnik**

U obrascima ljudskog govora postoje riječi koje se primjenjuju samo u određenom kontekstu. Međutim, kada se u sadržaju pojavljuju neobične ili neuobičajene riječi za takav kontekst, vjerojatno je generiran umjetnom inteligencijom. Također, odgovori mogu biti previše općeniti ili ne sasvim relevantni za postavljeno pitanje (McLean, 2024). Što se tiče gramatike i pravopisa, premda napredni UI-modeli mogu generirati tehnički savršenu gramatiku, u nekim slučajevima mogu postojati suptilne greške koje ljudi rijetko prave.

## **7.3. Predvidljivi uzorci**

Kada je sadržaj vrlo monoton i predvidljiv, čini se manje privlačnim nego onaj kojeg je napisala osoba koja želi zadržati pozornost čitatelja. Tekst djeluje poprilično formalno te može nedostajati prirodna varijacija koju stvarni autori imaju. Generirani tekst može izostaviti specifične detalje ili personalizirane informacije koje bi stručnjak ili autor s iskustvom uključio (McLean, 2024).

## **7.4. Duljina i struktura rečenice**

Kao zadnju vrstu koja odaje prisutnost umjetne inteligencije u kreiranju sadržaja potrebno je spomenuti nepromjenjivu duljinu i strukturu rečenice. Ljudi obično pišu raznolike rečenice kako bi bolje prikazali poruku koju žele prenijeti, dok UI-generatori uglavnom koriste ponavljajući obrazac rečeničnih struktura i duljina („How Does An AI Detector Work? And Is it Accurate?”, 2024).

Prepoznavanje UI-generiranog sadržaja zahtijeva pažljivo promatranje i korištenje dostupnih alata i metoda. Kombinacija tehnoloških rješenja i kritičkog razmišljanja ključna je za točnu identifikaciju. Također, potrebno je provjeriti vjerodostojnost izvora. Ako sadržaj dolazi iz nepoznatog ili sumnjivog izvora, postoji velika šansa da je takav tekst generiran umjetnom inteligencijom. Kako bi se dodatno potvrdila autentičnost, moguće je primijeniti i unakrsno provjeravanje informacija iz više izvora. Na taj način moguće je prepoznati nelogičnosti ili ponavljanja koja su karakteristična za UI-generiran sadržaj.

## 7.5. Primjer testiranja

U ovom odlomku bit će opisano testiranje navedenih metoda na primjeru riječi „drvar“. Prema definiciji koju je napisala osoba, riječ drvar jest imenica muškoga roda a-sklonidbe, a tvori se od imenice drvo te sufiksom -ar koji označava vršitelja radnje. Drvarom se smatra osoba koja trguje drvima te onaj koji siječe i prodaje drva (Hrvatski jezični portal, bez dat.). Iz ove definicije može se naslutiti da ju je napisala osoba jer su korištene jednostavne riječi i jasne rečenice za opisivanje tvorbe riječi. Osim toga, definicija riječi je točna i precizna, zajedno s gramatičkom klasifikacijom.



Slika 5. Objašnjenje riječi „drvar“

S druge strane, ChatGPT na pitanje „Kako nastaje riječ drvar?“, odgovara sljedeće:

*„Reč "drvar" nastaje na sličan način kao i reč "gljivar", kombinovanjem osnovne reči i sufiksa. U ovom slučaju:*

*1. Osnovna reč: drvo*

*2. Sufiks: -ar*

*Osnovna reč "drvo" odnosi se na materijal ili biljku, a sufiks "-ar" koristi se za formiranje imenica koje označavaju osobu koja je povezana sa određenom stvari ili delatnošću. Dakle, "drvar" je osoba koja se bavi sečenjem, sakupljanjem ili obradom drva.“*

*Spojivanjem ova dva elementa dobijamo reč "drvar", što znači osoba koja radi sa drvetom.*" (OpenAI, 2024).

Zaključno, prethodni odgovor ChatGPT-a na pitanje „Kako nastaje riječ drvar?“ sadrži nekoliko nepravilnosti. Prvo, u odgovoru je korištena riječ „spojivanje“, koja nije standardna u hrvatskom jeziku, umjesto pravilnog termina „tvorenje“ ili „nastajanje“. Drugo, odgovor je pisan na srpskom jeziku umjesto na hrvatskom, što uključuje upotrebu termina i gramatičkih konstrukcija koje nisu odgovarajuće za hrvatski standard. Ovi propusti ukazuju na potrebu za preciznijom jezičnom prilagodbom i korištenjem standardnih izraza specifičnih za hrvatski jezik prilikom objašnjavanja jezičnih fenomena. Na temelju prethodnog testiranja i dobivenih rezultata može se zaključiti da je odgovor na pitanje „Kako nastaje riječ drvar?“ generirala umjetna inteligencija, a ne čovjek.

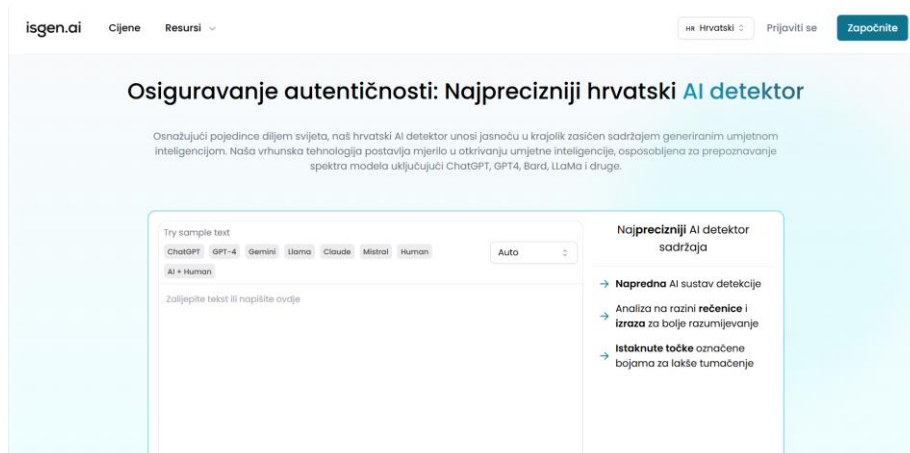
## 8. Detektori generiranih sadržaja

Alati za otkrivanje umjetne inteligencije su alati stvoreni za prepoznavanje i analizu sadržaja ili ponašanja generiranih sustava umjetne inteligencije. Ovi detektori mogu se koristiti u različitim domenama kako bi se osigurala autentičnost i integritet teksta, slike, zvuka i ponašanja. UI-detektor koristi tehnike obrade prirodnog jezika (NLP) i algoritme strojnog učenja za analizu specifičnih uzoraka u tekstu koji se obično prepoznaju kao pokazatelji sadržaja generiranog umjetnom inteligencijom. Također, oni služe i za otkrivanje plagijata te sprečavanje pristranih i neprikladnih UI-odgovora.

### 8.1. UI-detektori teksta

UI-detektori teksta identificiraju tekst generiran umjetnom inteligencijom. Koriste se kod otkrivanja plagijata, osiguravanja autentičnosti u objavljivanju te identificiranja objava na društvenim mrežama koje su generirali *botovi*.

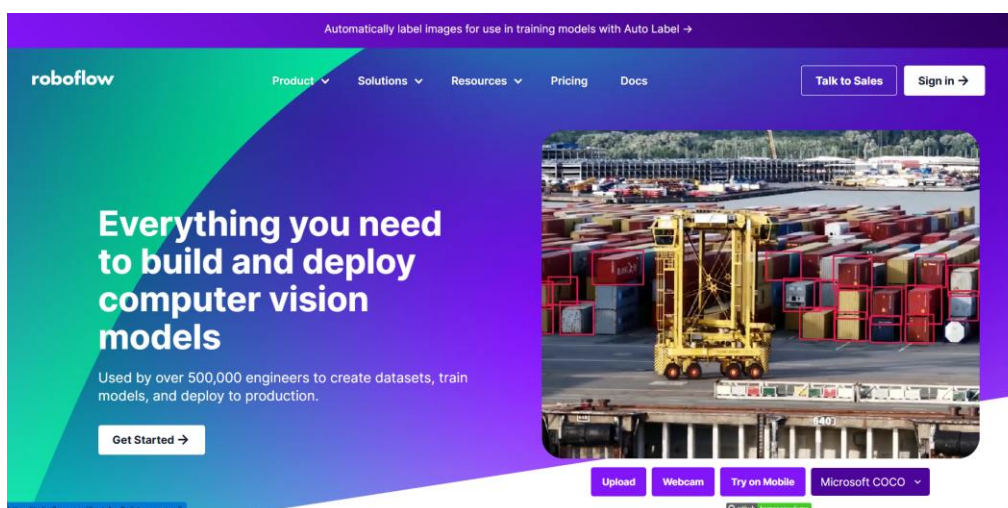
Postoje razni detektori za prepoznavanje izgeneriranog teksta, ali ipak valja spomenuti hrvatski UI-detektor pod nazivom Isgen.ai koji je osposobljen za prepoznavanje teksta napisanog umjetnom inteligencijom. Podržava više jezika, a posebno je koristan nastavnicima, studentima i kreatorima sadržaja. Platforma nudi nekoliko pretplatničkih planova, uključujući besplatnu opciju koja korisnicima omogućuje testiranje osnovnih mogućnosti detekcije UI-sadržaja bez ikakvih troškova. Isgenov sustav za detekciju umjetne inteligencije radi na način da analizira strukturu, odabir riječi i obrazaca u tekstu koristeći algoritme strojnog učenja trenirane na milijunima uzoraka. Može otkriti sadržaj generiran raznim UI-modelima, uključujući GPT-3.5, GPT-4, LLaMa i druge. Za one kojima su potrebne naprednije značajke ili druge korisne mogućnosti, Isgen.ai nudi planove do dvadeset tri dolara mjesečno koji uključuju prioritetnu podršku i pristup novim značajkama. Platforma također razvija dodatne alate poput API-ja za integraciju s drugim sustavima i detektor za UI-generirane videozapise i duboke lažnjake (*engl. deepfakeove*).



Slika 6. Isgen.ai

## 8.2. UI-detektori slike

UI-detektori slike razlikuju slike koje je stvorio čovjek od slika koje je generirala umjetna inteligencija. Provjeravaju autentičnost vizualnog sadržaja u medijima, otkrivaju duboke krivotvorine, osiguravaju integritet podataka te provjeravaju originalnost umjetničkih djela. Jedna od poznatih platformi za prepoznavanje objekata na slici jest Roboflow. Služi za pripremu podataka i treniranje modela za prepoznavanje te ima besplatnu verziju koja omogućuje osnovne funkcionalnosti provjere autentičnosti.

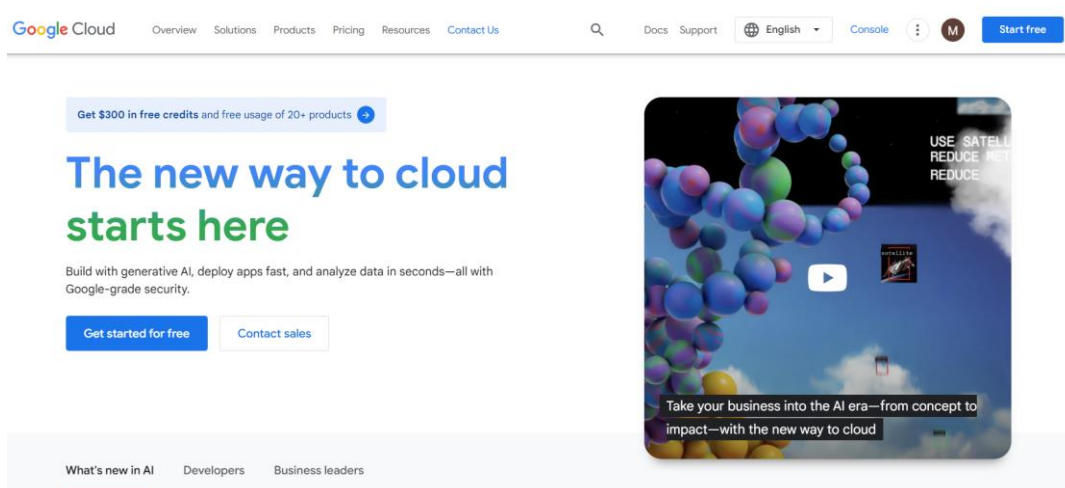


Slika 7. Roboflow



### 8.3. UI-detektori govora

UI-detektori govora otkrivaju govor ili zvuk generiran umjetnom inteligencijom. Pomažu kod identificiranja sintetičkih glasova u telekomunikacijama, provjere autentičnosti zvuka u medijima te otkrivanja dubokih krivotvorenja glasa. Ovi alati su ključni za osiguravanje integriteta audio materijala i sprječavanje manipulacija, što je posebno važno u kontekstu informacijske sigurnosti i vjerodostojnosti medijskog sadržaja. Poznat je Google Cloud Speech-to-Text API koji omogućuje prepoznavanje govora u realnom vremenu i podržava više jezika. Nudi visoku točnost i može se lako integrirati u *web* i mobilne aplikacije.

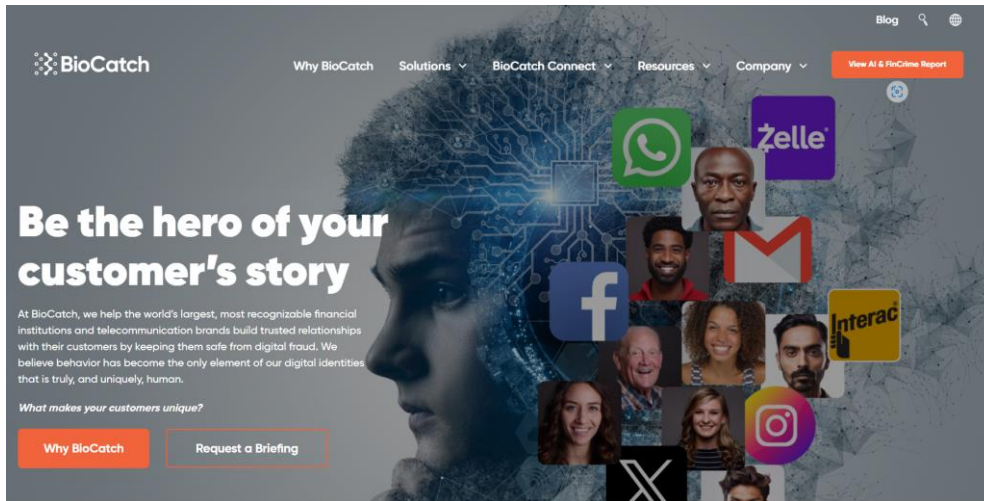


Slika 8. Google Cloud Speech-to-Text

### 8.4. Bihevioralni UI-detektori

Bihevioralni UI-detektori identificiraju obrasce ponašanja vođene umjetnom inteligencijom. Koriste se kod otkrivanja robota u igrama, prepoznavanja interakcija s korisnicima u e-trgovini koje se temelje na umjetnoj inteligenciji te uočavanja pokušaja hakiranja pomoću umjetne inteligencije u kibernetičkoj sigurnosti. Poznati bihevioralni UI-detektor jest BioCatch. Služi za kibernetičku sigurnost koji koristi bihevioralnu biometriku za prepoznavanje raznih prevara te u svrhe zaštite korisničkih naloga.

Analizira obrazce ponašanja kao što su dinamika pisanja po tipkovnici, pokreti miša i druge interakcije korisnika s aplikacijom.



Slika 9. BioCatch

Uz sve veću sofisticiranost UI-tehnologija, otkrivanje umjetne inteligencije postalo je ključno za održavanje povjerenja, autentičnosti i sigurnosti. Učinkovito prepoznavanje UI-sadržaja pomaže ublažiti rizike poput dezinformacija, prijevara i neovlaštene upotrebe umjetne inteligencije, osiguravajući integritet različitih vrsta sadržaja i interakcija u različitim domenama („Što su AI detektori?”, 2024).

## 9. Alati za detektiranje sadržaja generiranog umjetnom inteligencijom

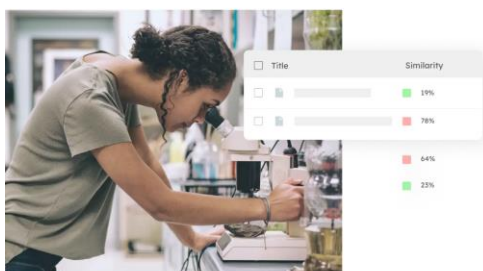
Moderni digitalni okoliš ne bi bio stabilan bez sustava za prepoznavanje UI-generiranog sadržaja koji nadziru automatsku distribuciju sadržaja na Internetu. Ovi sustavi primjenjuju najmodernije tehnike i metode strojnog učenja i obrade prirodnog jezika kako bi identificirali i klasificirali podatke generirane od strane UI-modela kao što su GPT-3, ChatGPT i slični. Oni imaju ključnu ulogu u sprečavanju širenja dezinformacija, zaštiti *online* zajednice te u prepoznavanju i uklanjanju lažnih vijesti.

Unatoč napretku u UI-tehnologiji, važnost alata za prepoznavanje teksta stvorenog umjetnom inteligencijom ostaje ključna za sigurnost korisnika, etički *online* diskurs i legalan *online* sadržaj. Ti alati igraju ključnu ulogu u očuvanju integriteta informacija, sprječavanju zloupotreba te promicanju odgovornog korištenja tehnologije. U ovom poglavlju ukratko su opisani alati za detekciju UI-generiranog teksta.

### 9.1. AI Text Detection API

AI Text Detection API omogućuje jednostavno lociranje i analiziranje teksta u različitim vrstama sadržaja. Ovo programsko sučelje aplikacije (*engl. Application Programming Interface* ili *API*) koristi najnoviju tehnologiju umjetne inteligencije kako bi se točno identificiralo i poslužilo tekstualnim sadržajem iz različitih izvora, uključujući fotografije, dokumente i digitalne medije. Njegova visoka preciznost i jednostavno sučelje omogućuju nastavnicima da prepoznaju plagijat u studentskim esejima i drugom UI-generiranom materijalu. Njegova jednostavna integracija u radne procese i upotreba od strane korisnika bez određenog tehničkog znanja smatra se velikom prednošću. Zahvaljujući moći obrade prirodnog jezika OpenAI-a, API može otkriti čak i blagi plagijat, osiguravajući jedinstvenost informacija. Zbog svoje sposobnosti analize teksta i pružanja detaljnog izvještaja, alat se može koristiti i za ocjenjivanje eseja te istraživanje strojnog učenja (Akram, 2023). Kao primjer, potrebno je spomenuti alat Turnitin, platformu za detekciju plagijarizma koja nudi usluge prepoznavanja UI-generiranog teksta. Turnitin može analizirati tekstove i identificirati obrasce karakteristične za UI-generirane sadržaje.

Student success starts here



Uphold academic integrity

Ensure original work from students and safeguard the value of writing.

Slika 10. Turnitin

## 9.2. ZeroGPT

ZeroGPT standard je za prepoznavanje dokumenata velikih jezičnih modela kao što je ChatGPT, a osmislio ga je Edward Tian. On identificira UI-sadržaj na razini fraze, odlomka i cijelog dokumenta, što ga čini iznimno prilagodljivim. Treniran je na raznolikom skupu ljudski napisanog i UI-generiranog teksta, dajući najveću pozornost engleskoj prozi. Korisnici mogu unijeti tekst za analizu korištenjem njegovog jednostavnog sučelja, a sustav pruža detaljne rezultate detekcije, ističući UI-detektiran materijal po rečenici kako bi se osigurala potpuna transparentnost. ZeroGPT podržava različite jezične modele umjetne inteligencije te koristi višeslojni pristup koji ima za cilj proizvesti što točnija predviđanja (Akram, 2023).

### Chat GPT detector by ZeroGPT: detect OpenAI text

ZeroGPT the most Advanced and Reliable Chat GPT detector tool

Despite these advancements, there are some potential drawbacks to using AI in SEO. One concern is the lack of transparency in how AI algorithms work and make decisions. This can make it difficult for SEO professionals to understand and trust the recommendations made by these tools.

Another potential issue is the risk of relying too heavily on AI and neglecting the importance of human expertise and creativity in SEO. While AI can provide valuable insights and suggestions, it is ultimately up to humans to create and implement effective SEO strategies.

In conclusion, AI is rapidly transforming the world of SEO, offering new tools and techniques for improving website visibility and ranking on search engine results pages. As AI continues to evolve, it is likely to play an increasingly important role in the future of SEO. However, it is important to approach AI-powered SEO tools with a critical eye and to balance their recommendations with human expertise and creativity.

Detect Text

Your text is AI/GPT Generated



Search Engine Optimization (SEO) is a crucial component of any digital marketing strategy.

The goal of SEO is to increase the visibility and ranking of a website on search engine results pages (SERPs) by optimizing its content, structure, and other factors.

With the rise of artificial intelligence (AI), SEO professionals have new tools and techniques at their disposal to achieve these objectives.

One area where AI is making a significant impact is in the analysis of user behavior.

Slika 11. ZeroGPT

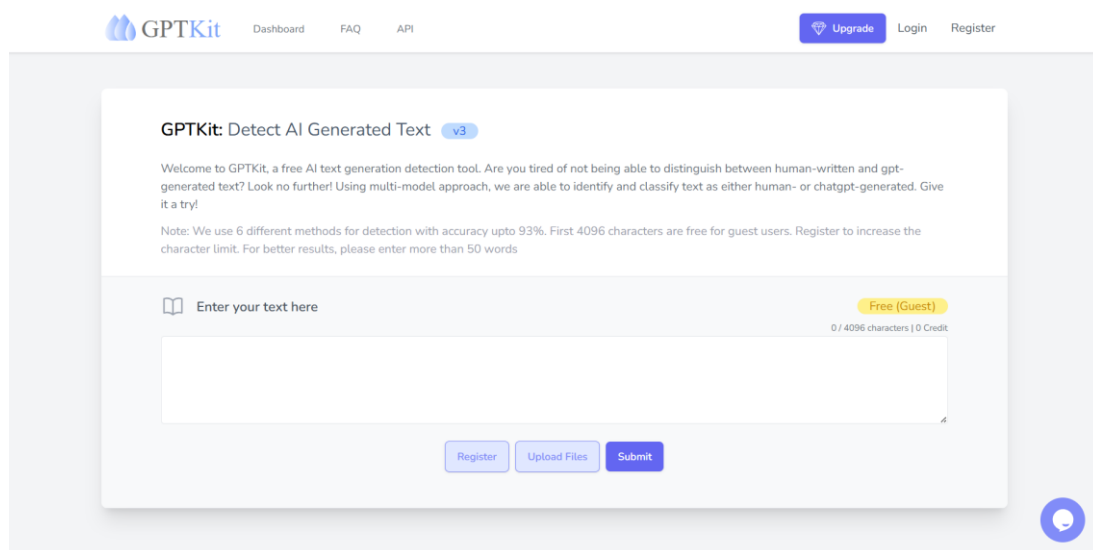
### 9.3. DeepAnalyse™ Technology

DeepAnalyse™ Technology model je za identifikaciju UI-sadržaja koji uključuje nekoliko komponenti koje analiziraju tekst radi utvrđivanja njegova podrijetla i prepoznavanja je li ga zaista napisala umjetna inteligencija. Koristi višestupanjsku metodologiju dizajniranu za optimizaciju preciznosti uz minimiziranje lažno pozitivnih i lažno negativnih rezultata. Model je specijaliziran za prepoznavanje sadržaja generiranog umjetnom inteligencijom kao što su ChatGPT, GPT-3, GPT-4, LLaMa i Gemini modeli. Konačno, primjenjuje sveobuhvatnu metodologiju dubokog učenja, treniranu na opsežnim zbirkama tekstova s Interneta, akademskim skupovima podataka te na korisničkim vlasničkim UI-skupovima podataka generiranih pomoću raznovrsnih jezičnih modela.

### 9.4. GPTKit

GPTKit razlikuje se od drugih alata po tome što koristi šest različitih metoda prepoznavanja sadržaja baziranih na umjetnoj inteligenciji, što značajno poboljšava preciznost u otkrivanju UI-generiranog materijala. Obuhvaća tehnike za osiguranje autentičnosti i točnosti sadržaja, uključujući provjeru činjenica u stvarnom vremenu, kombinaciju više izvora, kontekstualnu analizu, ljudsku verifikaciju, upotrebu

povijesnih podataka i strojno učenje na kvalitetnim podacima. Provjera činjenica u stvarnom vremenu upućuje na baze podataka kako bi se osigurala točnost informacija, dok kombinacija više izvora pomaže eliminirati pristranost. Kontekstualna analiza omogućava razumijevanje odnosa među informacijama, a ljudska verifikacija dodaje dodatni sloj sigurnosti za osjetljivije sadržaje gdje ljudi pregledavaju i potvrđuju točnost sadržaja prije objavljivanja. Korištenje povijesnih podataka osigurava da se generirani sadržaj temelji na već dokazanim informacijama, dok kontinuirano strojno učenje na kvalitetnim podacima poboljšava točnost i smanjuje mogućnost širenja dezinformacija. Zahvaljujući predanosti kontinuiranom istraživanju ovog alata, detektor u GPTKit-u sada postiže značajnu stopu točnosti od preko 93%, nakon što je treniran na širokom skupu podataka. GPTKit privremeno pohranjuje informacije radi obrade te ih nakon toga briše sa svog poslužitelja. Koristi se za provjeravanje autentičnih informacija te u obrazovne svrhe (Akram, 2023).

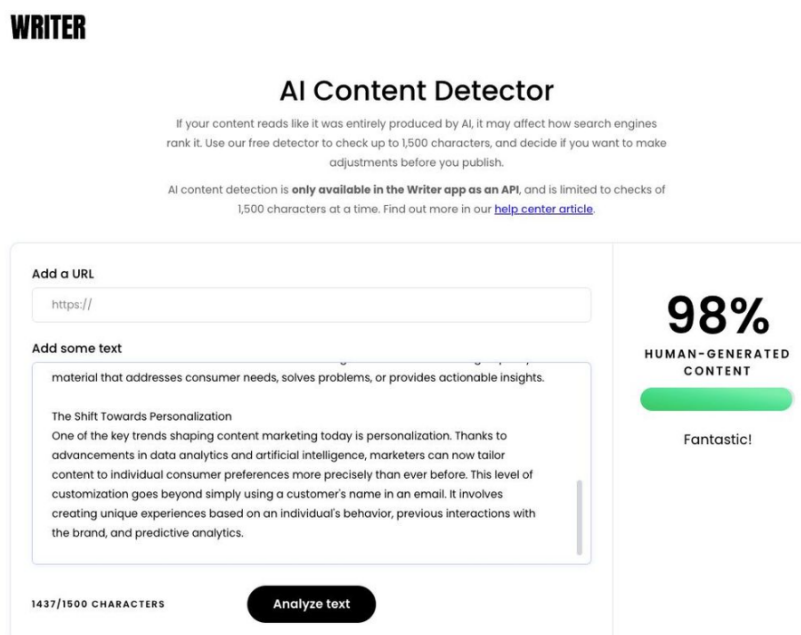


Slika 12. GPTKit

## 9.5. Writer's AI content detector

U početku kao alat za pomoć pri pisanju, Writer's AI content detector proširio je svoje usluge te se smatra detektorom sadržaja kreiranog umjetnom inteligencijom. Korištenjem napredne tehnike analize teksta i modela dubokog učenja, ovaj alat

pomaže u prepoznavanju tekstova generiranih umjetnom inteligencijom, s ciljem minimiziranja rizika povezanih s neautentičnim ili manipulativnim sadržajem. Koristi se za identifikaciju umjetne inteligencije u različitim izrazima, odlomcima i dokumentima. Zahvaljujući činjenici da je treniran na velikom skupu podataka, vrlo lako razlikuje tekst napisan umjetnom inteligencijom od onog kojeg je napisao čovjek. Dostupan je putem *web* stranice te sadrži i značajku za otkrivanje sadržaja na temelju URL-a (Amalia, 2023).

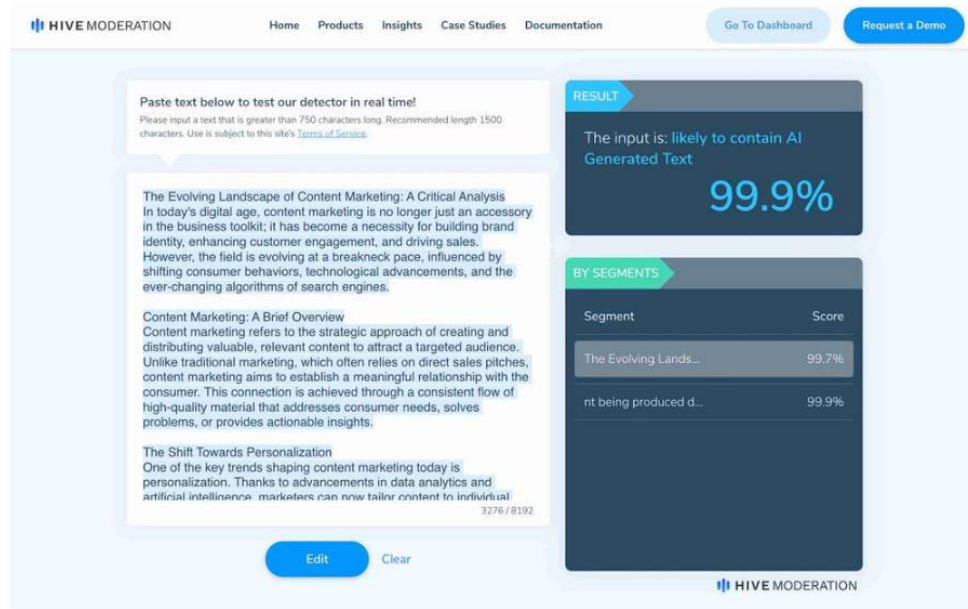


Slika 13. Writer's AI content detector

## 9.6. Hive Moderation AI content detector

Originalno usmjeren na moderiranje sadržaja, Hive Moderation AI content detector trenutno nudi mogućnost otkrivanja sadržaja generiranog umjetnom inteligencijom, što se može primijeniti na tekst, slike i audiozapise. Zanimljivost je da koristi napredne algoritme strojnog učenja i obrade prirodnog jezika za identifikaciju UI-generiranog sadržaja. Također, ključan je i u moderiranju internetskog sadržaja, sprječavajući širenje dezinformacija na način da štiti korisnika od neprimjerenog sadržaja i čuva integritet *online* zajednica. Osim toga, omogućuje učinkovitu i brzu analizu velike

količine materijala, što ga čini korisnim za platforme s visokim prometom ili mreže s velikim brojem korisnika (Amalia, 2023).

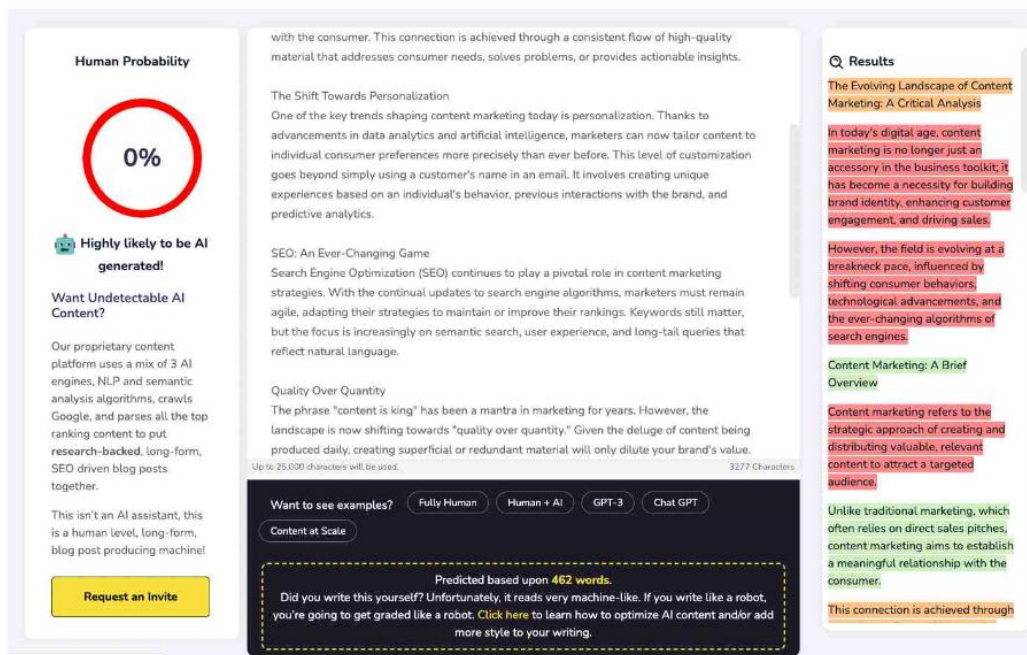


Slika 14. Hive Moderation AI content detector

## 9.7. Content at Scale AI Detector

Content at Scale AI Detector je alat koji koristi umjetnu inteligenciju kako bi odredio je li dio sadržaja generiran umjetnom inteligencijom ili je napisan od strane čovjeka. Koristi visoko precizne algoritme za otkrivanje UI-generiranog teksta obučene na milijardama pojedinačnih stranica podataka i riječi u više velikih jezičnih modela (LLM) radi predviđanja autentičnosti sadržaja (Amalia, 2023). Trenutačno je jedan od vodećih alata za prepoznavanje umjetne inteligencije u nekom sadržaju, a također je dizajniran i za identifikaciju karakteristika teksta generiranog umjetnom inteligencijom i ocjenjivanje sadržaja na temelju njegove predvidljivosti, vjerojatnosti i uzorka.



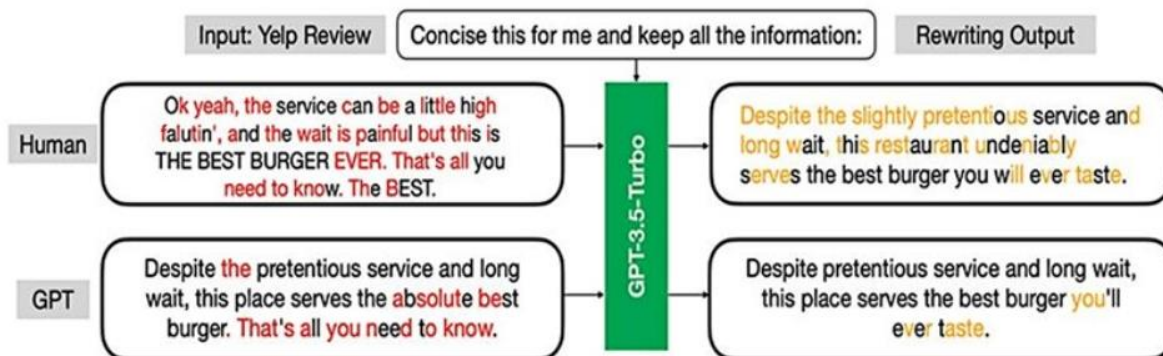


Slika 15. Content at Scale AI Detector

## 9.8. Raider

Znanstvenici računalstva s Columbia Engineeringa, Fakulteta za inženjerstvo i primijenjene znanosti na Sveučilištu Columbia u New Yorku, razvili su inovativnu metodu za prepoznavanje teksta generiranog umjetnom inteligencijom. Njihovo rješenje obećava revolucionirati provjeru autentičnosti digitalnog sadržaja, s ciljem smanjenja zabrinutosti u vezi s velikim jezičnim modelima, očuvanjem digitalnog integriteta, suzbijanjem dezinformacija te jačanjem povjerenja u nove tehnologije.

Raider, punim nazivom geneRative AI Detection via Rewriting, također je jedan od modela koji utvrđuje je li neki tekst napisao čovjek ili ga je pak generirala UI ili LLM, bez potrebe da se ulazi u unutarnji rad modela. On koristi jezični model za izmjenu zadanog teksta, a zatim mjeri koliko ga je sustav uređivao. Uzima dio teksta, poput objave na društvenim mrežama, recenzije proizvoda ili objave na blogu te od LLM-a traži da ga ponovno napiše. Nakon toga, uspoređuje izvorni tekst s iznova napisanim kako bi izmjerio izmjene. Velika količina izmjena znači da su tekst vjerojatno napisali ljudi, dok manje izmjena znači da je tekst najvjerojatnije strojno generiran (Smrekar, 2024).



Slika 16. Kako radi Raider?

## 9.9. Testiranje UI-detektora i njihovi rezultati

U svrhu istraživanja učinkovitosti alata za prepoznavanje umjetne inteligencije, potrebno je testirati nekoliko dostupnih UI-detektora. Cilj je provjeriti koliko točno navedeni alati mogu prepoznati tekstove generirane umjetnom inteligencijom u usporedbi s onima koje pišu ljudi. Testiranje će uključivati ZeroGPT, Hive Moderation AI content detector te Content at Scale AI detector kako bi se ocijenila njihova preciznost i pouzdanost u različitim scenarijima korištenja.

### 9.9.1. Testiranje ZeroGPT detektora

Testiranjem ZeroGPT alata, bilo je potrebno unijeti vlastiti tekst te nakon toga tekst koji je generirala umjetna inteligencija. Testirajući tekst koji u ovom radu objašnjava što je ZeroGPT, rezultat je pokazao da je 0% teksta generirano umjetnom inteligencijom, što potvrđuje točnost alata. ZeroGPT je precizno detektirao da je tekst autorski rad.

The screenshot displays the ZeroGPT web interface. At the top, there is a text input area containing a paragraph of text about GPTZero. Below the input area are two buttons: "Detect Text" and "Upload File". To the right of these buttons, it shows "579/15.000 Characters (Get up to 100,000 here)".

**Your Text is Human written**

0%  
AI GPT\*

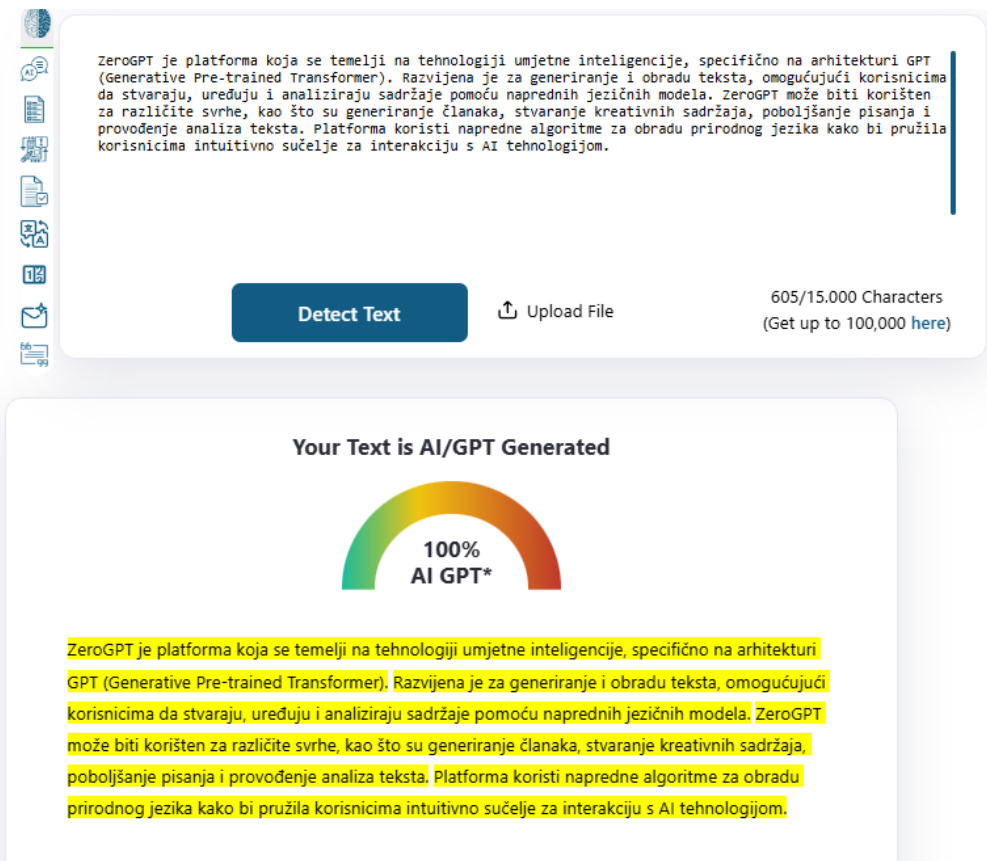
GPTZero standard je za prepoznavanje dokumenata velikih jezičnih modela kao što je ChatGPT, a osmislio ga je Edward Tian. On detektira UI sadržaj na razini fraze, odlomka i cijelog dokumenta, što ga čini vrlo prilagodljivim. ZeroGPT model treniran je na raznolikom skupu ljudski napisanog i UI-generiranog teksta, dajući najveću pozornost engleskoj prozi. Korisnici jednostavno mogu unijeti tekst za analizu korištenjem njegovo jednostavnog sučelja, a sustav pruža detaljne rezultate detekcije, naglašavajući UI-detektiran materijal po rečenici radi maksimalne transparentnosti.

■ Highlighted text is suspected to be most likely generated by AI\*  
579 Characters  
77 Words

Export to PDF

Slika 17. ZeroGPT-autorov tekst

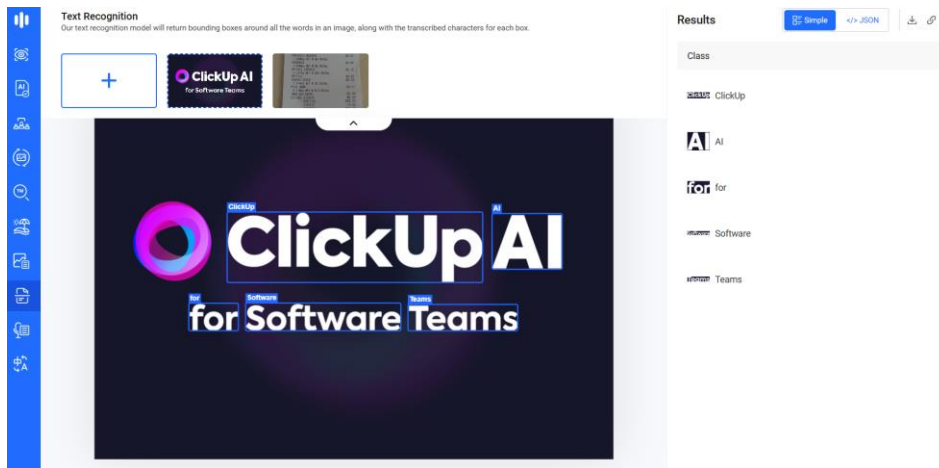
S druge strane, dajući zadatak ChatGPT-u da objasni što je ZeroGPT te ponovnim testiranjem alata na temelju izgeneriranog teksta, ispostavilo se da je ZeroGPT uspješno detektirao da je drugi tekst 100% rezultat umjetne inteligencije. Zaključno, alat se pokazao izuzetno korisnim i efikasnim jer precizno razlikuje autorski tekst od onog kojeg je izgenerirala umjetna inteligencija.



Slika 18. ZeroGPT-tekst UI

### 9.9.2. Testiranje Hive Moderation AI content detektora

Testirajući Hive moderation alat, analizirane su dvije slike različite složenosti kako bi se evaluirala njegova sposobnost prepoznavanja teksta. Prva slika je bila jednostavnija, s malo teksta i jasno definiranim elementima, dok je druga slika bila složenija, s raznolikim rasporedom riječi različitih veličina. U oba slučaja, Hive je uspješno identificirao tekstualni sadržaj bez obzira na kompleksnost slike. Ovaj test potvrđuje visoku pouzdanost i efikasnost Hive moderation alata u razumijevanju i obradi vizualnog sadržaja, što ga čini korisnim alatom za analizu sadržaja.



Slika 19. Hive Moderation-jednostavna fotografija

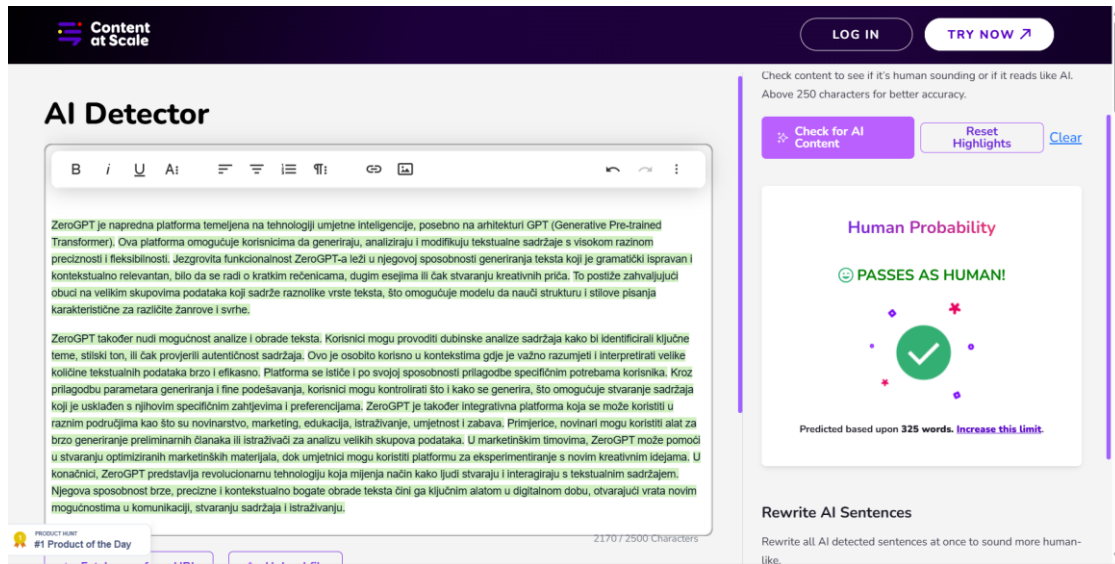


Slika 20. Hive Moderation-složena fotografija

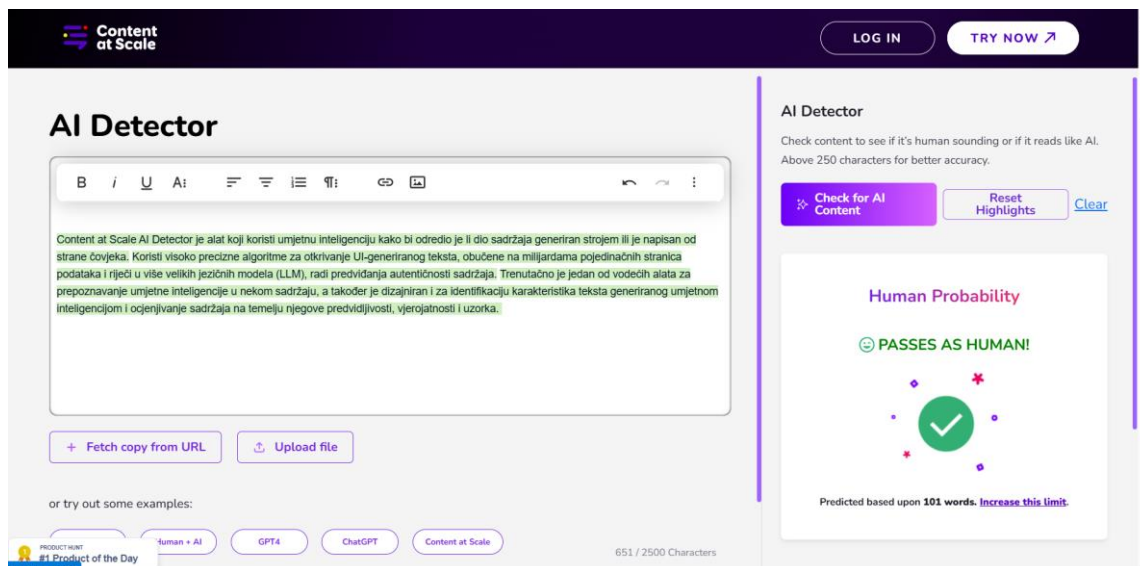
### 9.9.3. Testiranje Content at Scale AI detektora

Testiranjem Content at Scale AI detector alata na dva teksta od kojih je jedan napisan od strane čovjeka, a drugi izgeneriran umjetnom inteligencijom, otkrivena je značajna nedosljednost u rezultatima. U oba slučaja, detektor je klasificirao oba teksta kao ljudski napisana, što dovodi u pitanje točnost alata. Ovi rezultati potiču sumnje u pouzdanost UI-alata za detektiranje autentičnosti sadržaja, osobito s napredovanjem generativnih modela koji proizvode tekst koji je teško razlikovati od ljudskog pisanja.

Analitička procjena ovakvih alata postaje ključna za njihovu praktičnu primjenu, zahtijevajući daljnje istraživanje i poboljšanja kako bi se osigurala pouzdanost u identifikaciji izvora teksta.

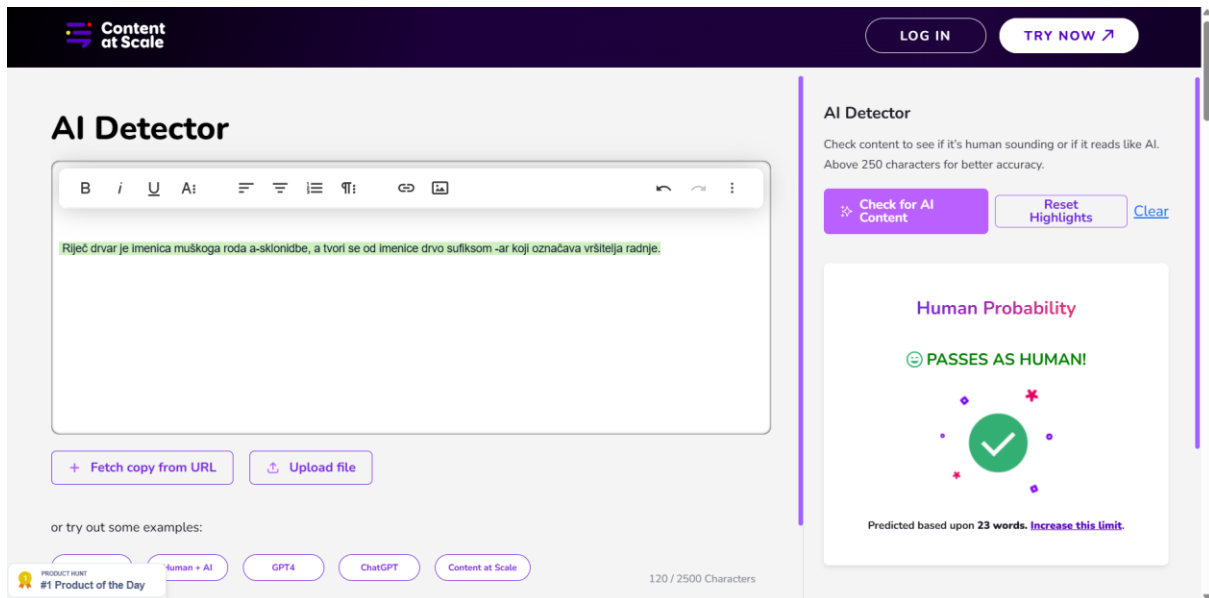


Slika 21. Content at Scale AI detector-generiran tekst



Slika 22. Content at Scale AI detector-autorov tekst

Kao dodatna potvrda da alat nije u potpunosti pouzdan, iskorišten je primjer riječi „drvar” kao i kod testiranja u poglavlju gdje je objašnjeno kako prepoznati UI-generiran sadržaj. U prvom slučaju je testiran ljudski tekst kojeg je alat točno prepoznao, odnosno označio ga je tekстом pisanim od strane čovjeka.



Slika 23. Ljudski tekst

Nakon što je alat točno prepoznao da je prvi tekst napisan od strane čovjeka, unesen je tekst koji je generirao ChatGPT. Međutim, alat je i ovaj tekst prepoznao kao ljudski. Ovo testiranje pokazalo je da alat može pogrešno klasificirati tekstove te ih označiti istima, iako oni ne dolaze iz istog izvora. To dodatno potvrđuje činjenicu da trenutna tehnologija za detekciju umjetne inteligencije nije savršena i može imati poteškoće u razlikovanju između ljudskog i UI-generiranog sadržaja.

**Content at Scale** LOG IN TRY NOW ↗

## AI Detector

Check content to see if it's human sounding or if it reads like AI. Above 250 characters for better accuracy.

**Check for AI Content** **Reset Highlights** [Clear](#)

**Human Probability**

😊 **PASSES AS HUMAN!**

Predicted based upon **87 words**. [Increase this limit.](#)

or try out some examples:

PRODUCT HUNT #1 Product of the Day human + AI GPT4 ChatGPT Content at Scale 498 / 2500 Characters

Rec "drvar" nastaje na sličan način kao i rec "gijlvar", kombinovanjem osnovne reci i sufiksa. U ovom slučaju:

1. "Osnovna rec": drvo
2. "Sufiks": -ar

Osnovna rec "drvo" odnosi se na materijal ili biljku, a sufiks "-ar" koristi se za formiranje imenica koje označavaju osobu koja je povezana sa određenom stvari ili delatnošću. Dakle, "drvar" je osoba koja se bavi sečenjem, sakupljanjem ili obradom drva.

Spojivanjem ova dva elementa dobijamo rec "drvar", što znači osoba koja radi sa drvetom.

+ Fetch copy from URL Upload file

Slika 24. ChatGPT-izgeneriran tekst

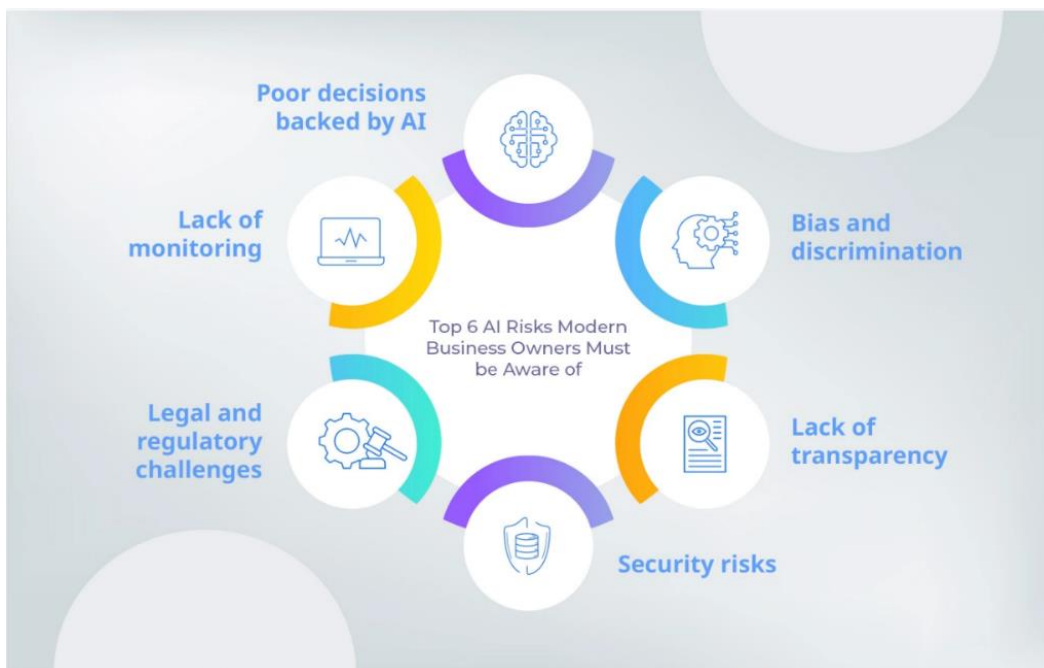


## 10. Rizici korištenja umjetne inteligencije

Nedavnim napretkom u umjetnoj inteligenciji, posebno u području generativnosti, unaprijed obučeni transformatorski (GPT) veliki jezični modeli (LLM) doveli su do niza javno dostupnih *online* alata za generiranje teksta. Budući da su ovi modeli uvježbani na tekstovima koje su napisali ljudi, sadržaj generiran ovim alatima može biti prilično teško razlikovati od ljudski napisanih tekstova. Veliki jezični modeli često se susreću s izazovima u provjeravanju točnosti informacija koje generiraju. Na taj način se povećava mogućnost kruženja netočnih i pogrešnih informacija. Među najvećim rizicima upotrebe generativne UI-tehnologije su oni vezani za intelektualno vlasništvo. Generativna UI-tehnologija koristi neuronske mreže koje se mogu obučavati na velikim postojećim skupovima podataka kako bi stvorile nove podatke ili objekte kao što su tekst, slika, audio ili videozapisi. Korištenjem generativne UI-tehnologije omogućena je brzina i efikasnost različitim poduzećima, ali ona također dovodi i do snažnih iskušenja koje zaposlenici mogu vrlo lako zloupotrijebiti. Na primjer, radnici bi mogli lagati da su odradili određeni posao, a zapravo bi ga umjesto njih odradila umjetna inteligencija. Na taj način, bili bi dobro plaćeni za zadatke koje zapravo nisu ni obavili („The flip side of generative AI: Challenges and risks around responsible use”, 2023). Osim toga, alati umjetne inteligencije ne ažuriraju se u stvarnom vremenu te nisu u skladu s trenutnim pravnim okvirima. Time stvaraju potencijalne rizike u područjima kao što je autorsko pravo. Prepoznavanje sadržaja generiranog umjetnom inteligencijom može pomoći u smanjenju rizika i ojačati vjerodostojnost kreiranih sadržaja s kojima korisnici svakodnevno dolaze u kontakt putem Interneta („The Risks of AI-Generated Content: What You Need To Know“, 2023).

Umjetna inteligencija donosi brojne prednosti, ali isto tako nosi i određene rizike koje treba pažljivo razmotriti. Prvo, nedostatak transparentnosti u nekim UI-sustavima, osobito onima baziranim na dubokom učenju, može rezultirati teškoćama u objašnjavanju odluka koje takvi sustavi donose. Osim toga, privatnost i sigurnost podataka su ključni problemi, budući da UI-sustavi mogu biti meta napada i zlouporabe, što može ugroziti osobne podatke i osjetljive informacije korisnika. Dalje, postoji rizik od autonomnog djelovanja umjetne inteligencije bez adekvatnog ljudskog nadzora, što dovodi do nepredviđenih posljedica i pogrešaka. Socijalni i ekonomski

aspekti također igraju važnu ulogu s potencijalom za povećanje postojećih nejednakosti i ekonomskih promjena, uključujući gubitak radnih mjesta zbog automatizacije. Na kraju, regulacija i etički okviri za upotrebu UI-tehnologija su ključni za minimiziranje rizika i negativnih utjecaja na društvo i gospodarstvo. Sve ove aspekte treba pažljivo balansirati kako bi se iskoristile prednosti umjetne inteligencije uz smanjenje potencijalnih rizika i negativnih posljedica za društvo (Isik, Joshi i Goutas, 2024).



Slika 25. Rizici umjetne inteligencije

## 11. Kako označiti sadržaj kojeg je generirala umjetna inteligencija?

Američki zastupnik Ritchie Torres, ove je godine predložio Zakon o otkrivanju umjetne inteligencije koji zahtijeva odricanje od odgovornosti i označivanje sadržaja kojeg je ona kreirala, što uključuje fotografije, tekst, videozapise i audiozapise. Takav sadržaj bi trebao biti označen riječima „ovaj sadržaj je generirala umjetna inteligencija”. S druge strane, Koalicija za porijeklo i autentičnost sadržaja (C2PA) je razvila tehnički standard za praćenje podrijetla sadržaja kako bi utvrdila je li takvim sadržajem na bilo koji način manipulirano. Pod Koaliciju se ubrajaju razne organizacije kao što su Microsoft, Intel, BBC, Sony i drugi. Vodeni žigovi i *disclaimer* također se smatraju dobrim načinom označivanja sadržaja, a vremenom se pokazalo kako korisnici društvenih mreža ne dijele i ne vjeruju sadržajima koji su označeni kao dezinformacija. Također, korisnici moraju znati da se radi o umjetnoj inteligenciji ukoliko se stalno ponavljaju žargoni ili pak zbunjujući i besmisleni izrazi (Smrekar, 2023).

Što se tiče UI-oznake, ona ima više prednosti. Njena zadaća je da označi da je sadržaj generiran uz pomoć umjetne inteligencije te da upozori da isti takav sadržaj korisnike može dovesti u zabludu koja će ih uputiti na krivo mjesto i pogrešno informirati. Smatra se da bi uvođenje različitih oznaka za lažan sadržaj kojeg je kreirala umjetna inteligencija bio jednako dobar korak kao i za oznake sadržaja koji nije generiran od strane umjetne inteligencije. Javlja se problem i kod generiranja samih oznaka. Kod samooznačivanja, mali broj kreatora bi stvarno priznao da je njihov sadržaj s namjerom lažan. Stoga se javljaju opcije kao što su strojno učenje, nabava iz mnoštva (*engl. crowdsourcing*) te digitalna forenzika.

Konačno, istraživači su otkrili da neke oznake neće funkcionirati svugdje. Na primjer, govornici kineskog povezivali su riječ *artificial* s ljudskom uključenošću, dok taj izraz na engleskom, portugalskom i španjolskom označava automatizaciju. Istraživači su zaključili da se u različitim zemljama koriste različite oznake kako bi se osiguralo da pojmovi budu jednako jasni svima.

U Europi, uporaba umjetne inteligencije uskoro će biti regulirana prvom sveobuhvatnom zakonodavnom inicijativom pod nazivom Akt o umjetnoj inteligenciji.

EU planira osigurati bolje uvjete za razvoj i upotrebu umjetne inteligencije te drugih novih tehnologija. Komisija je u travnju 2021. predložila prvi regulatorni okvir EU-a za umjetnu inteligenciju, koji uključuje analizu i klasifikaciju sustava umjetne inteligencije prema različitim razinama rizika koje predstavljaju za korisnike. Različite razine rizika određivat će stupanj regulacije koji će biti primijenjen („Europski parlament”, 2023).

## 12. Zaključak

Očigledno je da ljudska uključenost ostaje najbitnija u izradi sadržaja. Ipak, naveliko su dostupni različiti alati i smjernice za uviđanje teksta generiranog umjetnom inteligencijom. Identifikacija teksta generiranog umjetnom inteligencijom postala je ključna u eri digitalnih komunikacija i masovnog stvaranja sadržaja. Premda su UI-sustavi izuzetno napredovali i sposobni su generirati visokokvalitetne tekstove koji često mogu zavarati čitatelje, postoje suptilne razlike koje omogućuju njihovo prepoznavanje. Analizom stila pisanja, kontekstualne dosljednosti, metapodataka i ponašanja moguće je razlikovati UI-generiran sadržaj od onoga kojeg su stvorili ljudi. Međutim, uz napredak u alatima za generiranje sadržaja razvijeni su i sofisticirani alati za otkrivanje generiranih tekstova koji koriste napredne algoritme i metode kako bi identificirali obrasce tipične za UI-generiran tekst. Oni pomažu u očuvanju autentičnosti i integriteta informacija, omogućujući korisnicima da razlikuju ljudski napisane tekstove od onih generiranih umjetnom inteligencijom.

Unatoč impresivnim sposobnostima umjetne inteligencije, ljudi ostaju jedini pravi autori koji unose jedinstvenu kreativnost, emocionalnu dubinu i intuitivno razumijevanje konteksta u svoje pisanje. Ljudski autori posjeduju sposobnost prilagodbe i inovacije koje UI još uvijek ne može u potpunosti replicirati. Dok UI može služiti kao moćan alat za podršku i unapređenje stvaranja sadržaja, ljudski doprinos ostaje nezamjenjiv u stvaranju autentičnih i značajnih tekstova. U konačnici, kombinacija ljudske kreativnosti i tehnološke podrške može dovesti do najboljih rezultata, ali prepoznavanje i vrednovanje ljudskog autorstva ostaje ključno za očuvanje autentičnosti i kvalitete informacija u digitalnom dobu.

## 13. Literatura

### KNJIGE

Russel, S. (2022). *Kao čovjek: umjetna inteligencija - napredak ili prijetnja?*.

Swarup, P. (2012). *Artificial Intelligence*.

### MREŽNI IZVORI, WEB STRANICE

Akram, A. (2023). *An Empirical Study of AI-Generated Text Detection Tools*.

Preuzeto 19. travnja 2024. s

[https://www.researchgate.net/publication/375225294\\_An\\_Empirical\\_Study\\_of\\_AI-Generated\\_Text\\_Detection\\_Tools](https://www.researchgate.net/publication/375225294_An_Empirical_Study_of_AI-Generated_Text_Detection_Tools)

Awan, A. A. (2023). *What is Text Generation?*. Preuzeto 3. svibnja 2024. s

<https://www.datacamp.com/blog/what-is-text-generation>

Amalia, A. (2023). *How to identify AI-generated text: With or without software*.

Preuzeto 18. travnja 2024. s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Bennett, M. (2024). *Artificial intelligence vs. human intelligence: Differences explained*. Preuzeto 23. travnja 2024. s

<https://www.techtarget.com/searchenterpriseai/tip/Artificial-intelligence-vs-human-intelligence-How-are-they-different>

Bretous, M. (2023). *How Jasper.ai Writing Assistant Works. Its Pros and Cons*.

Preuzeto 24. svibnja 2024. s <https://blog.hubspot.com/marketing/jasper-ai>

Baldikov, N. (2024). *The Conclusive Writesonic Review for 2024: Pros, Cons, and Everything in Between*. Preuzeto 12. lipnja 2024. s

<https://inboundblogging.com/writesonic-review/>

Chaudhary, S. (2023). *Pros and Cons of Jasper AI for Content Writing*. Preuzeto

12. lipnja 2024. s <https://writerarmy.com/pros-and-cons-of-jasper-ai-for-content-writing>

- Drvar. (bez dat.) U Hrvatski jezični portal. Preuzeto 17. lipnja 2024. s [https://hjp.znanje.hr/index.php?show=search\\_by\\_id&id=fF9hXhg%3D&key\\_word=drvar](https://hjp.znanje.hr/index.php?show=search_by_id&id=fF9hXhg%3D&key_word=drvar)
- Europski parlament (2023). *Akt EU-a o umjetnoj inteligenciji: Prva regulacija tog područja*. Preuzeto 13. travnja 2024. s <https://www.europarl.europa.eu/topics/hr/article/20230601STO93804/akt-eu-a-o-umjetnoj-inteligenciji-prva-regulacija-tog-podrucja>
- How Does An AI Detector Work? And Is it Accurate? (2024). Preuzeto 5. svibnja 2024. s <https://smodin.io/hr/blog/how-ai-detectors-work-accurate/#:~:text=Op%C4%87enito%2C%20alati%20poput%20ovog%20kriste%20nekoliko%20va%C5%BEnih%20metoda,s%20kojima%20su%20ve%C4%87%20upoznati.%20...%20Vi%C5%A1e%20stavki>
- Isik, Ö., Joshi, A., i Goutas, L. (2024). *4 Types of Gen AI Risk and How to Mitigate Them*. Preuzeto 5. svibnja 2024. s <https://hbr.org/2024/05/4-types-of-gen-ai-risk-and-how-to-mitigate-them>
- Kako točno rade AI detektori? (2023). Preuzeto 5. svibnja 2024. s [https://blog.plag.ai/hr/how-exactly-do-ai-detectors-work#google\\_vignette](https://blog.plag.ai/hr/how-exactly-do-ai-detectors-work#google_vignette)
- Manuel, B. (2024). *What Is ClickUp & Why Use the Project Management Software in 2024?*. Preuzeto 16. svibnja 2024. s <https://www.cloudwards.net/what-is-clickup/#:~:text=In%20short%2C%20ClickUp%20is%20a%20cloud-based%20software%20and,what%20they%20do%20will%20benefit%20from%20ClickUp%E2%80%99s%20offerings.>
- McLean, D. (2024). *How to Detect AI Writing in 2024 (Expert Tips)*. Preuzeto 7. svibnja 2024. s <https://www.elegantthemes.com/blog/business/how-to-detect-ai-writing>
- OpenAI. (2024). ChatGPT (GPT-4) [Generirani tekst]. <https://www.openai.com/>
- Smrekar, M. (2024). *Raider, lektor koji prepoznaje AI generirane tekstove*. Preuzeto 18. travnja 2023. s <https://www.bug.hr/umjetna-inteligencija/raidar-lektor-koji-prepoznaje-ai-generirane-tekstove-39480>

Što su AI detektori? (2024). Preuzeto 28. travnja 2024. s <https://textflip.ai/hr/ai-content-detector/>

The Risks of AI-Generated Content: What You Need To Know (2023). Preuzeto 15. svibnja 2024. s <https://www.upwork.com/resources/ai-generated-content-risks>

The flip side of generative AI: Challenges and risks around responsible use (2023). Preuzeto 13. svibnja 2024. s <https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/generativeai-risks.pdf>

Ward, J. (2023). *How to Detect AI Content: Proven Methods for Distinguishing Human from Machine Writing*. Preuzeto 13. svibnja 2024. s <https://www.linkedin.com/pulse/how-detect-ai-content-proven-methods-distinguishing-human-ward-6q74e/>



## **Popis oznaka i kratica**

*AI*-Artificial Intelligence

*API*- Application Programming Interface

*C2PA*- Koalicija za porijeklo i autentičnost sadržaja

*EU*-Europska unija

*GPT*-Generative Pre-trained Transformer

*IT*-Information technology

*LLM*-Large Language Model

*NLP*- Neuro-linguistic programming

*PaLM*- Pathways Language Model

*SEO*- Search Engine Optimization

*UI*-Umjetna inteligencija

*URL*- Uniform Resource Locator

## Popis slika

Slika 1. Ljudska ili umjetna inteligencija?

Preuzeto s <https://www.consilium.europa.eu/hr/documents-publications/library/library-blog/posts/artificial-intelligence-and-facial-recognition-talk-with-leading-ai-experts/>

Slika 2. Jasper A.I.

Preuzeto s [https://www.jasper.ai/?adgroupid=163111846019&campaignid=21338182930&utm\\_source=google&utm\\_term=&utm\\_content=&utm\\_campaign=mm\\_search\\_brand\\_na&utm\\_medium=cpc&hsa\\_acc=5024544109&hsa\\_cam=21338182930&hsa\\_grp=163111846019&hsa\\_ad=700760558871&hsa\\_src=g&hsa\\_tgt=kwd-1547008706560&hsa\\_kw=jasper%20ai&hsa\\_mt=e&hsa\\_net=adwords&hsa\\_ver=3&gad\\_source=1&gclid=Cj0KCQjwvb-zBhCmARIsAAfUI2uYeAGLHSOJfWHAbp6hG6CjIDF\\_8chA0GU2FUyVK3kqEx0G-\\_H140aAuOqEALw\\_wcB](https://www.jasper.ai/?adgroupid=163111846019&campaignid=21338182930&utm_source=google&utm_term=&utm_content=&utm_campaign=mm_search_brand_na&utm_medium=cpc&hsa_acc=5024544109&hsa_cam=21338182930&hsa_grp=163111846019&hsa_ad=700760558871&hsa_src=g&hsa_tgt=kwd-1547008706560&hsa_kw=jasper%20ai&hsa_mt=e&hsa_net=adwords&hsa_ver=3&gad_source=1&gclid=Cj0KCQjwvb-zBhCmARIsAAfUI2uYeAGLHSOJfWHAbp6hG6CjIDF_8chA0GU2FUyVK3kqEx0G-_H140aAuOqEALw_wcB)

Slika 3. Writesonic

Preuzeto s <https://writesonic.com/features>

Slika 4. ClickUp AI

Preuzeto s <https://clickup.com/ai>

Slika 5. Objašnjenje riječi „drvar”

Preuzeto s [https://hjp.znanje.hr/index.php?show=search\\_by\\_id&id=fF9hXhg%3D](https://hjp.znanje.hr/index.php?show=search_by_id&id=fF9hXhg%3D)

Slika 6. Isgen.ai

Preuzeto s <https://isgen.ai/hr>

Slika 7. RoboFlow

Preuzeto s <https://roboflow.com/>

Slika 8. Google Cloud Speech-to-Text

Preuzeto s [https://cloud.google.com/speech-to-text/?utm\\_source=bing&utm\\_medium=cpc&utm\\_campaign=emea-emea-all-en-dr-bkws-all-all-trial-e-gcp-1011340&utm\\_content=text-ad-none-any-DEV\\_c-CRE\\_-ADGP\\_Hybrid+%7C+BKWS+-+EXA+%7C+Txt+-+AI+And+Machine+Learning+-+Text+to+Speech+-+v1-KWID\\_43700078980688107-kwd-77309753369888:loc-49-userloc\\_138911&utm\\_term=KW\\_google%20speech%20to%20text%20online-NET\\_o-PLAC\\_&&refclickid=NEE\\_HotelDigitalMediaCampaign&pmed=DPM\\_CVG\\_TRV\\_HO\\_T\\_NNE\\_SEM\\_GOG\\_&qclid=0936017473281caba29267041454ad64&qclsrc=3p.ds&msclkid=0936017473281caba29267041454ad64](https://cloud.google.com/speech-to-text/?utm_source=bing&utm_medium=cpc&utm_campaign=emea-emea-all-en-dr-bkws-all-all-trial-e-gcp-1011340&utm_content=text-ad-none-any-DEV_c-CRE_-ADGP_Hybrid+%7C+BKWS+-+EXA+%7C+Txt+-+AI+And+Machine+Learning+-+Text+to+Speech+-+v1-KWID_43700078980688107-kwd-77309753369888:loc-49-userloc_138911&utm_term=KW_google%20speech%20to%20text%20online-NET_o-PLAC_&&refclickid=NEE_HotelDigitalMediaCampaign&pmed=DPM_CVG_TRV_HO_T_NNE_SEM_GOG_&qclid=0936017473281caba29267041454ad64&qclsrc=3p.ds&msclkid=0936017473281caba29267041454ad64)

Slika 9. BioCatch

Preuzeto s <https://www.biocatch.com/>

Slika 10. Turnitin

Preuzeto s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Slika 11. ZeroGPT

Preuzeto s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Slika 12. GPTKit

Preuzeto s <https://gptkit.ai/>

Slika 13. Writer's AI content detector

Preuzeto s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Slika 14. Hive Moderation AI content detector

Preuzeto s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Slika 15. Content at Scale AI Detector

Preuzeto s <https://www.contentgrip.com/how-to-spot-ai-generated-text/>

Slika 16. Kako radi Raider?

Preuzeto s <https://www.bug.hr/umjetna-inteligencija/raidar-lektor-koji-prepoznaje-ai-generirane-tekstove-39480>

Slika 17. ZeroGPT-autorov tekst

Preuzeto s <https://www.zerogpt.com/>

Slika 18. ZeroGPT-tekst UI

Preuzeto s <https://www.zerogpt.com/>

Slika 19. Hive Moderation-jednostavna fotografija

Preuzeto s <https://hivemoderation.com/>

Slika 20. Hive Moderation-složena fotografija

Preuzeto s <https://hivemoderation.com/>

Slika 21. Content at Scale AI detector-generiran tekst

Preuzeto s <https://contentatscale.ai/ai-content-detector/>

Slika 22. Content at Scale AI detector-autorov tekst

Preuzeto s <https://contentatscale.ai/ai-content-detector/>

Slika 23. Ljudski tekst

Preuzeto s <https://contentatscale.ai/ai-content-detector/>

Slika 24. ChatGPT-izgeneriran tekst

Preuzeto s <https://contentatscale.ai/ai-content-detector/>

Slika 25. Rizici umjetne inteligencije

Preuzeto s <https://binmile.com/blog/potential-ai-risks-in-business/>

# Identifikacija teksta generiranog umjetnom inteligencijom

## Sažetak

Svrha rada je bila istražiti identifikaciju teksta generiranog umjetnom inteligencijom (UI) te analizirati ručne i softverske metode za otkrivanje takvog sadržaja. Kroz rad je opisano što identifikacija teksta generiranog umjetnom inteligencijom podrazumijeva, istaknute su ručne metode identifikacije koje su temeljene na ljudskoj procjeni i iskustvu te softverske metode koje se oslanjaju na algoritme i tehnologije strojnog učenja.

Analizirale su se prednosti i mane različitih UI-generatora sadržaja poput Jasper A.I.-ja, Writesonica, Clickupa i dr. istaknuvši njihovu učinkovitost i potencijalne probleme kao što su nepouzdanost ili nedostatak autentičnosti. Također, rad je istražio najpoznatije alate koji se koriste za otkrivanje UI-generiranog sadržaja kao što su ZeroGPT, Writer's AI content detector, Hive Moderation AI content detector i dr., analizirajući njihovu točnost i pouzdanost u identifikaciji takvog sadržaja.

Na kraju, u radu su se istražile mogućnosti označavanja sadržaja generiranog umjetnom inteligencijom te se razmotrilo pitanje o tome kako pravilno označiti takav sadržaj kako bi se korisnicima pružila transparentnost i pouzdanost. Cilj je bio pružiti pregled postupaka i alata koji se koriste za identifikaciju UI-generiranog teksta, istaknuti rizike takvog sadržaja te analizirati najpoznatije UI-generatore sadržaja i alate za otkrivanje UI-generiranog sadržaja kako bi korisnicima pružili bolje razumijevanje i svijest o mogućnostima i izazovima povezanim s korištenjem UI-tehnologija u stvaranju i prepoznavanju sadržaja.

**Ključne riječi:** umjetna inteligencija, UI-generator, UI-detektor, identifikacija, generirani sadržaj

# Identification of text generated by Artificial Intelligence

## Summary

The purpose of the paper was to investigate the identification of text generated by Artificial Intelligence (AI) and to analyze manual and software methods for detecting such content. The paper describes what the identification of text generated by artificial intelligence entails, highlights manual methods of identification that are based on human judgment and experience, and software methods that rely on algorithms and machine learning technologies.

The advantages and disadvantages of various AI content generators such as Jasper A.I., Writesonic, Clickup, etc. were analyzed, highlighting their effectiveness and potential problems such as unreliability or lack of authenticity. Also, the paper investigated the most famous tools used to detect AI-generated content such as ZeroGPT, Writer's AI content detector, Hive Moderation AI content detector, etc., analyzing their accuracy and reliability in identifying such content.

To conclude, this paper explores the possibilities of tagging content generated by Artificial Intelligence and considers the question of how to properly label such content in order to provide transparency and reliability to users. The goal was to provide an overview of the procedures and tools used to identify AI-generated text, highlight the risks of such content and analyze the most popular AI content generators and AI-generated content detection tools to provide users with a better understanding and awareness of the opportunities and challenges associated with using AI-technologies in creating and recognizing content.

**Key words:** Artificial Intelligence, AI-generator, AI-detector, identification, generated content