



Sveučilište u Rijeci  
University of Rijeka  
<http://www.uniri.hr>

*Polytechnica: Journal of Technology Education, Volume 7, Number 2 (2023)*  
*Politehnika: Časopis za tehnički odgoj i obrazovanje, Svezak 7, Broj 2 (2023)*



Politehnika  
Polytechnica  
<https://politehnika.uniri.hr>  
[cte@uniri.hr](mailto:cte@uniri.hr)

DOI: <https://doi.org/10.36978/cte.7.2.1>

Izvorni znanstveni članak  
Original scientific paper  
UDK: 004.8:37

# Izazovi korištenja jezičnih modela OpenAI platforme u kontekstu vrednovanja studentskih rješenja projektnih zadataka iz domene računalnog programiranja

Danijel Kučak

Visoko učilište Algebra  
Studij primijenjenog računarstva  
Gradišćanska 24, 10000 Zagreb  
[danijel.kucak@algebra.hr](mailto:danijel.kucak@algebra.hr)

## Sažetak

Autor ovog istraživanja godinama svjedoči tome kako je motivacija studenata za kontinuirani rad relativno niska. Kako bi povećali motivaciju studenata za samostalni rad, nastavnici na Visokom učilištu Algebra odlučili su posljednjih godina promijeniti pristup. Umjesto inzistiranja da studenti uče za ispit u određeno vrijeme, dobili su niz projekata koje su trebali samostalno implementirati. Rezultati istraživanja pokazali su da postoji napredak u kasnije ostvarenim rezultatima u odnosu na tradicionalne ispite. S druge strane, pojavom moćnih i besplatnih ili vrlo pristupačnih jezičnih modela kao što su oni platforme OpenAI, nastavnici se susreću sa izazovima jer studenti mogu koristiti takve modele kako bi riješili projektni zadatak, što je u suprotnosti s etičkim kodeksom. Svrha ovoga istraživanja je utvrditi kakve rezultate je sposoban postići jezični model platforme OpenAI, ako mu se zada projektni zadatak iz određenoga kolegija. Definirana su 2 istraživačka pitanja. Kakav rezultat postiže jezični model na OpenAI platformi kada mu se zada projektni zadatak iz kolegija Java web programiranje i postoji li razlika u rezultatima u odnosu na postignute rezultate studenata (Q1)? Postoji li razlika u rezultatima koje postiže jezični model u odnosu na prirodni jezik na kojem mu je zadan projektni zadatak (hrvatski ili engleski) (Q2)? Analiza rezultata pokazala je da je u određenim slučajevima jezični model sposoban riješiti ispit za prolaznu ocjenu bez dodatnih intervencija korisnika, iako su postignuti rezultati značajno slabiji od rezultata koje postižu studenti. Bolje rezultate jezični model postiže ako mu se ispit zada na engleskom jeziku u odnosu na rezultate na hrvatskom jeziku.

**Ključne riječi:** vrednovanje znanja; učenje bazirano na projektima; OpenAI; akademski integritet; umjetna inteligencija.

## 1 Uvod

Kao dugogodišnji nastavnik na visokoškolskim ustanovama, autor ovoga rada konstantno ima izazov kreirati sadržaj i pristup nastavnom procesu koji će maksimalno motivirati studente na rad. Kako bi se poboljšala postignuća studenata na sveučilištu, često je potrebno koristiti netradicionalne metode i

postupke poučavanja (Močinić, 2012). Uobičajena organizacija obrazovnog procesa i filozofija usmjerena na konvergentno mišljenje kako postoji „samo jedan točan odgovor“ često kočite kreativnost, umjesto da ju razvijaju (Bognar, 2012). Stoga je bitno da nastava bude moderna i kreativna kako bi se povećala motivacija za učenje i sudjelovanje studenata u nastavnom procesu (Debeljak De Martini, Ristić,

Potočnjak, 2020). Suvremeni obrazovni programi trebali bi započeti s nastavnicima, koji bi preuzeli ulogu refleksivnih praktičara i akcijskih istraživača kako bi stvorili promjene za poboljšanje obrazovnog sustava iznutra prema van (Zawacki-Richter, Marin, Bond, Gouvernour, 2019).

Navedeno predstavlja izazov i motivacijski faktor autoru rada i studentima. Generacije današnjih studenata trebale bi se u učenju voditi rješavanjem problema različitim metodama istraživanja u kontroliranim uvjetima, uz nadzor nastavnika, kako bi postali aktivni sudionici, a ne pasivni promatrači (Vinduška, 2016). Primjerice, suvremene nastavne metode poput istraživanja i interaktivne nastave, koje predstavljaju odmak od pristupa "ex-cathedra" i usmjerene su na aktivno sudjelovanje studenata i samostalnu primjenu znanja (Cigan, Šlogar, 2012), dokazano stvaraju veći interes kod studenata. Također, budući da klasičan pristup studiranju stvara veliki stres za studente, „nastavnici bi trebali podržati autonomiju studenata i olakšati im akademski tijek kako bi spriječili izgaranje studenata“ (Ljubin-Golub, Rijavec, 2020).

Jedan od pristupa koji treba poduzeti je korištenje studija slučaja koje su najprije oblikovane nastavnikovim pogledom na predmet, a kasnije su izgrađene na temelju očekivanja, pozadine i sposobnosti učenika koje pomažu u razvoju sposobnosti analize i razmišljanja. Kao drugi primjer, 'dizajn razmišljanje' uvodi praktično rješavanje problema koje studenta stavlja u središte problema, često dovodeći do diferencijacije, pa čak i inovacije (Pap, Vdović, Baletić, 2019). Nadalje, primjena načela igrifikacije u nastavi pokazala je vrlo dobre rezultate (Kučak, Bele, Pašić, 2021) jer "igrifikacija koristi mehaniku igre za transformaciju iskustva učenja u igru, dok učenje temeljeno na igricama integrira igru u proces učenja" (Kučak, Pašić, Mršić, 2022). Naposljetku, mnoga su istraživanja dokazala da projektna nastava potiče aktivnost i zadovoljstvo učenika što može povećati uspjeh studenata (Kerić i sur., 2017). To možda još više vrijedi za kolegije iz područja računalnog programiranja jer se studenti boje apstraktne prirode programiranja. Često je teško shvatiti svrhu stečenog znanja (Konecki, Lovrenčić, Kaniški, 2016). Na temelju prethodno navedenih činjenica, postojeći obrazovni sustav zahtijeva značajne promjene kako bi se učenici pripremili za buduće radne izazove. Provjera stečenog znanja trebala bi se više usmjeriti na kreativno izražavanje učenika, a manje na reprodukciju sadržaja (Dubovicki, 2019).

Postoje razni pristupi u metodologiji poučavanja i evaluacije stečenog znanja (Mohorovičić, Strčić, 2011). Jedan od smjerova koji zastupaju nastavnici su projektna provjere znanja. Umjesto tradicionalnog ispita koji se piše u kontroliranim uvjetima unutar

zadanog vremenskog roka, studenti dobivaju projektni zadatak na kojem rade samostalno u nekontroliranim uvjetima. Projekti su u pravilu složeniji od tradicionalnih ispita, ali su znatno realističniji i bliži stvarnim izazovima s kojima će se studenti kasnije susretati u radnom okruženju. Traže veći angažman studenata u pronalaženju rješenja, pa samim time i omogućuju kvalitetnije stjecanje znanja. Kako bi se osigurala provjera autentičnosti studentskih projekata, nastavnici najčešće provode usmene verifikacije predanih studentskih projektnih rješenja. Postoji nekoliko studija koje su pokazale da takav pristup donosi pozitivne pomake u stečenom znanju studenata u odnosu na tradicionalne evaluacije ispitom u kontroliranim uvjetima (Gratchev, Yeng, 2018) i (Yuliansyah, Ayu, 2021). Također, postoje i recentna istraživanja u Hrvatskoj koja to potvrđuju

. Autori istraživanja (Bele, Kučak, Pašić, 2023) pokazali su na studiji slučaja provedenoj na Visokom učilištu Algebra da postoje pozitivni pomaci u rezultatima i potrebnom vremenu polaganja kolegija koji su projektno orijentirani, odnosno imaju projektni zadatak kao model evaluacije stečenog znanja, nasuprot tradicionalnih ispita.

Izazovi bi, u tom pristupu učenja baziranom na projektnom radu, mogli biti od prije nekoliko godina široko dostupni i trenutno često besplatni ili vrlo jeftini jezični modeli bazirani na umjetnoj inteligenciji. Takvi modeli imaju puno mogućnosti, a jedna od njih je da su sposobni kreirati programski kod na osnovu upita definiranih prirodnim jezikom. Postoji nekoliko takvih jezičnih modela, a najpoznatiji i najpopularniji su jezični modeli bazirani na platformi kompanije OpenAI (OpenAI Platform, 2023).

Upotreba jezičnih modela kao što su oni tvrtke OpenAI i sličnih, izazvao je zanimanje u akademskoj zajednici, odnosno među istraživačima i nastavnicima (Biswas, 2023). Jezični modeli obrade prirodnog jezika koji je razvio OpenAI koriste veliki skup podataka za generiranje tekstualnih odgovora na upite učenika, povratne informacije i upute (Gilson i sur., 2023). Mogu simulirati razgovore s učenicima, davati povratne informacije, odgovarati na pitanja i pružati podrške, kao i kreirati programski kod na osnovu upita definiranih prirodnim jezikom. Godine 2021. tvrtka OpenAI predstavila je svoj jezični model Codex. Jezični model Codex potomak je jednog od najnaprednijih tada dostupnih modela prirodnog jezika GPT-32. Codex je istreniran na više od 50 milijuna repozitorija na platformi Github. Codex može uzeti upute na engleskom ili bilo kojem drugom jeziku i sposoban je generirati programski kod u različitim programskim jezicima. Također, može, između ostalog, prevoditi kod iz jednoga u drugi programski jezik i objasniti na engleskome jeziku funkcionalnost dostavljenoga programskoga koda. Codex je u ožujku 2023. godine zamijenjen još naprednijim jezičnim modelima iz

familije gpt-3-5-turbo i gpt-4. Korisnicima nudi internet sučelje za unos upita, a omogućuje i korištenje programskog sučelja za integraciju jezičnih modela u korisnička programska rješenja (OpenAI Platform, 2023).

Jezični modeli imaju potencijal pomoći studentima u procesu učenja. Međutim, s druge strane, brza implementacija ovih jezičnih modela, kao što je Chat GPT tvrtke OpenAI ili Bard tvrtke Google, također predstavlja nekoliko izazova. U ovom članku, fokusirat ćemo se na izazov mogućeg utjecaja korištenja takvih jezičnih modela na proces evaluacije znanja, odnosno akademskog integriteta. Autori recentnih istraživanja su pokazali da taj izazov postoji i da se treba suočiti s njim (Perkins, 2023), (Sullivan, Kelly, McLaughlan, 2023), (Mijwil i dr., 2023) i (Finnie-Ansley i dr., 2022).

## 2 Metodologija istraživanja

Kako je navedeno u uvodu, postoji mnogo različitih pristupa u korištenju nastavnih metoda iz kolegija vezanih za računalno programiranje, pa to za sobom povlači i različite modele evaluacije stečenoga znanja studenata. Jedna od metoda su i kolegiji koji stečeno znanje studenata provjeravaju kroz izradu projektnog zadatka. Cilj ove studije je istražiti utjecaj korištenja jezičnih modela u rješavanju projektnih zadataka iz domene računalnog programiranja. Studenti studija primijenjenog računarstva na Visokom učilištu Algebra, smjer programsko inženjerstvo, na trećoj godini preddiplomskog studija upisuju izborni kolegij Java web programiranje. U akademskoj godini 2021/2022 kolegij je upisalo 68 studenata. Studenti su podijeljeni u 2 grupe, jedna nastavu sluša na engleskom jeziku, druga na hrvatskom, i to je jedina razlika u nastavnom procesu. Isti nastavnik predaje objema grupama. Studenti su početkom kolegija dobili projektni zadatak koji nastavnici kreiraju za svaku akademsku godinu, što znači da studenti nisu mogli projekt vidjeti prije. Grupa koja je slušala na engleskom jeziku, dobila je projektni zadatak na engleskom jeziku, dok je grupa koja je slušala na hrvatskom jeziku, dobila zadatak na hrvatskom jeziku. Nema razlike u projektnom zadatku za obje grupe, osim u jeziku na kojem je tekst zadatka napisan. Gradivo kolegija podijeljeno je na 6 ishoda učenja, što se reflektira i na projektnom zadatku. Specifični zahtjevi projektnog zadatka koji se tiču konkretnih ishoda učenja, označeni su oznakama I1 do I6. Ishodi učenja definirani su na početku nastavnog procesa i objavljeni studenti ma kroz sustav digitalne referade. Struktura projekta za danog na hrvatskom jeziku može se vidjeti na slici 1.

Kreirajte Spring aplikaciju koja omogućava Internet kupnju odabranih proizvoda. Aplikacija ima sljedeće karakteristike/svojstva:

- Postoje 2 role u aplikaciji: korisnik (kupac) i administrator

### KORISNIK

- Anonimni korisnik može pregledavati kategorije (**anonymous**)
- Anonimni korisnik može pregledavati proizvode (**anonymous**)
- Korisnik može dodati proizvod u košaricu (**anonymous**)
- Korisnik može definirati količinu određenog proizvoda kojeg stavlja u košaricu (npr. 20 zaštitnih folija za mobilni telefon) (**anonymous**)
- Korisnik može mijenjati sastav košarice (npr. skroz izbaciti neki proizvod ili određenu količinu nekog proizvoda iz košarice ili sve proizvode) (**anonymous**)
- Korisnik može izvršiti online kupnju proizvoda, moguće su opcije gotovina - pouzeće i PayPal (sve informacije oko implementacije PayPal-a, kreiranja sandbox accounta itd, možete pronaći na <https://developer.paypal.com/>) (**authenticated - kupac**)
- Korisnik mora vidjeti povijesni pregled **svjkih** kupnji sa sljedećim podacima: (**authenticated - kupac**)
  - Što je kupljeno (sadržaj košarice, sa količinama)
  - Kada je kupljeno
  - Kako je kupljeno (gotovina – pouzeće ili PayPal)

### ADMINISTRATOR

- Administrator upravlja kategorijama i proizvodima (CRUD operacije nad kategorijama i proizvodima) (**authenticated - admin**)
- Administrator može vidjeti povijesni pregled podataka o tome tko se, kada i sa koje adrese ulogira (**authenticated - admin**)
- Administrator može vidjeti kompletan povijesni pregled kupnji, za sve kupce, s istim informacijama. Omogućiti jednostavne filtere za pretragu po kupcu i vremenskom periodu kupnje (**authenticated - admin**)

### OPĆE UPUTE

- Obavezno koristiti Spring MVC + Thymeleaf
- Implementirati autentifikacijske/autorizacijske elemente aplikacije prema zadanoj specifikaciji korištenjem Spring Security pristupa

Slika 1. Projektni zadatak zadan studentima

Za potrebe objektivnog i nepristranog ispravljanja ispita, na kolegiju je definirana shema ocjenjivanja, na osnovu koje ispravljači evaluiraju studentske projekte. Na Slici 2 su prikazani kriterija ocjenjivanja za 2 ishoda učenja, LO3 i LO4 (engl. learning outcome, skraćeno LO).

| Aspect Description (LO3 - max 10 points)   | Partial Marking Instructions or Judgement Score Description   | Max Mark |
|--|---|----------|
| Spring MVC is properly used to handle request scope of the view data model.                                      | Deduct half a point if the request scope is not properly used.  | 1,00     |
| Spring MVC is properly used to handle session scope of the view data model.                                      | Deduct half a point if the session scope is not properly used.  | 1,00     |
| Spring MVC annotated validation components are properly used.  | Deduct half a point if the validation components are not properly used.   | 3,00     |
| Tags for displaying validation messages are properly used.   | Deduct half a point if the validation messages are not properly used.   | 1,00     |
| Spring MVC session management is properly implemented (created and invalidated).                                 | Deduct half a point if the Spring MVC session management is not properly implemented.   | 1,00     |
| AJAX components are properly used for implementation of asynchronous communication.                              | Deduct a point of AJAX components are not properly used. Deduct all points if AJAX components are not used at all.  | 3,00     |
| Aspect Description (LO4 - max 10 points)   | Partial Marking Instructions or Judgement Score Description   | Max Mark |
| Spring security configuration class is properly annotated by appropriate annotations with deployment descriptor. | Deduct half a point to Spring Security configuration class is not properly annotated.   | 2,00     |
| Multiple user accounts with roles are properly defined.  | Deduct half a point if no multiple user accounts are defined.   | 1,00     |
| Spring Security authentication features (login and logout) are properly implemented with deployment descriptor.  | Deduct a point if the login feature is not properly implemented.<br>Deduct half a point if logout feature is not properly implemented.<br>Deduct all points if not login and logout functionality are implemented at all. | 3,00     |
| View components are properly enhanced with security tags that implement authorization features.                  | Deduct half a point if security tags are not properly implemented. Deduct all points if no security tags are used.  | 1,00     |
| Spring Security authentication features for multiple user roles are properly configured.                         | Deduct a point if authentication features for multiple user roles are not implemented properly.   | 2,00     |
| Spring Security authorization features are properly implemented.   | Deduct half a point if authentication features for multiple user roles are not implemented properly.  | 1,00     |

Slika 2. Dio sheme ocjenjivanja projektnog zadatka

Maksimalni broj bodova koje student može ostvariti za rješenje projektnog zadatka je 100 bodova. Pri tome je bitno naglasiti kako student iz svakog od 6 ishoda učenja mora imati minimalno 50% ostvarenih bodova za taj ishod učenja. Studenti su ispit mogli

polagati na 2 ispitna termina u siječnju 2022. godine. Ispitu je pristupilo ukupno 55 studenata, te su nakon primjene svih opisanih kriterija ostvarili rezultate prikazane u Tablici 1.

| Ocjena                     | Nedovoljan (1) | Dovoljan (2) | Dobar (3) | Vrlo dobar (4) | Odličan (5) |
|----------------------------|----------------|--------------|-----------|----------------|-------------|
| Broj studenata (ukupno 55) | 10             | 12           | 8         | 10             | 15          |

Tablica 1. Distribucija ocjena studenata

Cilj ovog istraživanja je utvrditi kakve rezultate će postići jezični model OpenAI kada mu se zada isti projektni zadatak koji su dobili i studenti. Zanima nas može li položiti ispit prema utvrđenim kriterijima identičnima kao i za studente, kakav rezultat je postigao u usporedbi sa studentima i postoji li razlika u njegovim rezultatima, ako mu se zada projekt na hrvatskome u odnosu na engleski jezik. U tu svrhu, postavljena su sljedeća istraživačka pitanja:

- Q1: Kakav rezultat postiže jezični model na OpenAI platformi kad mu se zada projektni zadatak iz kolegija Java web programiranje i postoji li razlika u rezultatima u odnosu na postignute rezultate studenata?
- Q2: Postoji li razlika u rezultatima koje postiže jezični model na OpenAI platformi kad mu se zada projektni zadatak iz kolegija Java web programiranje u odnosu na prirodni jezik na kojem mu je zadan projektni zadatak (hrvatski ili engleski)?

Za potrebe ovog istraživanja, koristio se OpenAI jezični model daVinci-003. Preko web sučelja na platformi openai.com zadan je projektni zadatak na hrvatskom i engleskom jeziku. Radi prirode samog web sučelja, originalni projektni zadatak neznatno je modificiran radi jednostavnijeg unosa na web sučelje platforme, te je u potpunosti unesen na sljedeći način:

- Kreirajte Spring aplikaciju koja omogućava Internet kupnju odabranih proizvoda. Aplikacija ima sljedeće karakteristike/svojstva:
- Postoje 2 role u aplikaciji: korisnik (kupac) i administrator
- Anonimni korisnik može pregledavati kategorije
- Anonimni korisnik može pregledavati proizvode
- Anonimni korisnik može dodati proizvod u košaricu
- Anonimni korisnik može definirati količinu određenoga proizvoda kojega stavlja u košaricu (npr. 20 zaštitnih folija za mobilni telefon)
- Anonimni korisnik može mijenjati sastav košarice (npr. skroz izbaciti neki proizvod ili

određenu količinu nekog proizvoda iz košarice ili sve proizvode)

- Prijavljeni korisnik može izvršiti online kupnju proizvoda; moguće su opcije gotovina - pouzeće i PayPal
- Prijavljeni korisnik mora vidjeti povijesni pregled svojih kupnji sa sljedećim podacima:
  - Što je kupljeno (sadržaj košarice, sa količinama)
  - Kada je kupljeno
  - Kako je kupljeno (gotovina – pouzeće ili PayPal)
- Prijavljeni administrator upravlja kategorijama i proizvodima (CRUD operacije nad kategorijama i proizvodima)
- Prijavljeni administrator može vidjeti povijesni pregled podataka o tome tko se, kada i sa koje adrese prijavio
- Prijavljeni administrator može vidjeti kompletan povijesni pregled kupnji, za sve kupce, s istim informacijama.
- Omogućiti jednostavne filtere za pretragu po kupcu i vremenskom periodu kupnje
- Obavezno koristiti Spring MVC + Thymeleaf
- Implementirati autentifikacijske/autorizacijske elemente aplikacije prema za danoj specifikaciji korištenjem Spring Security pristupa

Jezični model je u prvom odgovoru postavio kostur projekta po koracima i dao osnovne implementacije određenih koraka. Slika 3 prikazuje primjer dijela generiranog odgovora na zadani unos.

Svako zadavanje zadatka po svakom jeziku je ponovljeno 20 puta. Primjer jednog malog dijela generiranog odgovora prikazan je na slici 3.

```

1. Postavljanje projekta:
Počnite s postavljanjem Spring Boot projekta. Koristite Spring Initializr ili Maven za generiranje projekta s potrebnim ovisnostima, uključujući Spring Web, Spring Data JPA, Thymeleaf, Spring Security, i H2 bazu podataka (za potrebe ovog primjera).

2. Kreiranje entiteta:
Definirajte klase za entitete kao što su 'Kategorija', 'Proizvod', 'Kupac', 'Košarica', 'Kupnja' i 'Administrator'. Ovisno o vašim potrebama, dodajte potrebne attribute, anotacije i relacije među njima.

3. Konfiguracija baze podataka:
Konfigurirajte svoju bazu podataka putem 'application.properties' ili 'application.yml'. Npr., za H2 bazu podataka:

spring:
  datasource:
    url: jdbc:h2:mem:testdb
    driverClassName: org.h2.Driver
    username: sa
    password:
  h2:
    console:
      enabled: true

4. Implementacija sigurnosti:
Konfigurirajte Spring Security kako biste podržali autorizaciju i autentifikaciju. Postavite različite uloge (korisnik, administrator) i odgovarajuće ovlasti (permissions) za pristup određenim resursima. Na primjer:

```

Slika 3. Primjer dijela generiranog programskog koda

Nakon dobivenog kostura projekta, za svaki korak definiran odgovorom od strane jezičnog modela, zadan je novi unos na osnovu rezultata dobivenih u koraku ranije. Na primjer, za korak 2 prikazan na slici 3, zadan je unos jezičnom modelu: „Definirajte klase za entitete kao što su Kategorija, Proizvod, Kupac, Košarica, Kupnja i Administrator“. Na slici 4 može se vidjeti primjer generirane klase Proizvod.

```
@Entity
public class Proizvod {
    no usages
    @Id
    @GeneratedValue(strategy = GenerationType.IDENTITY)
    private Long id;
    no usages
    private String naziv;
    no usages
    private String opis;
    no usages
    private double cijena;
    no usages
    private int kolicina;

    no usages
    @ManyToOne
    private Kategorija kategorija;

    // Getters and setters
}
```

Slika 4: Primjer generirane klase Proizvod

Tako generirani artefakti dodani su u projektno rješenje i cjelokupno rješenje se bodovalo prema unaprijed zadanoj shemi ocjenjivanja.

### 3 Rezultati

Projektni zadatak, prikazan na slici 1, na način opisan u prethodnom poglavlju, zadan je jezičnom modelu na platformi OpenAI po 20 puta na hrvatskom i engleskom jeziku. Kako jezični model ne daje uvijek za iste ulazne podatke iste izlazne podatke, na taj način je autor htio dobiti relevantniji uzorak. Nakon svakog zadavanja projektnog zadatka, izvršena je evaluacija rezultata prema unaprijed utvrđenim kriterijima. Dobiveni rezultati prikazan je u tablici 2.

| Ocjena   | Nedovoljan (1) | Dovoljan (2) | Dobar (3) | Vrlo dobar (4) | Odličan (5) |
|--|----------------|--------------|-----------|----------------|-------------|
| Broj polaganja na hrvatskom jeziku (ukupno 20) | 12             | 8            | 0         | 0              | 0           |
| Broj polaganja na engleskom jeziku (ukupno 20) | 9              | 10           | 1         | 0              | 0           |

Tablica 2: distribucija ocjena jezičnog modela za ENG i HRV

### 3.1 Kakav rezultat postiže jezični model na OpenAI platformi kad mu se zada projektni zadatak iz kolegija Java web programiranje i postoji li razlika u rezultatima u odnosu na postignute rezultate studenata?

Kada je projektni zadatak bio zadan na hrvatskom jeziku 20 puta, nakon evaluacije projekata, 12 ponavljanja dobilo je nedovoljnu ocjenu, 8 ponavljanja dobilo je dovoljnu ocjenu, dok niti jedno ponavljanje nije dobilo ocjenu veću od dovoljan. Ako usporedimo prosječnu ocjenu studenata na promatranom uzorku (3.15), sa prosječnom ocjenom koju je dobio jezični model (1.4 za projekt zadan na hrvatskom jeziku, te 1.6 za projekt zadan na engleskom jeziku), možemo zaključiti kako je rezultat studenata značajno bolji u odnosu na rezultat koji je postigao jezični model. Korištenjem  $\chi^2$  testa autor rada htio je metodološki potvrditi da postoji i statistički značajna razlika između dobivenih ocjena studenata i jezičnog modela. Distribucije ocjena za studente i jezični model, kad mu je projekt zadan na hrvatskom jeziku, prikazan je u tablici 3.

| Ocjena            | Nedovoljan (1) | Dovoljan (2) | Dobar (3)  | Vrlo dobar (4) | Odličan (5) | Ukupno |
|-------------------|----------------|--------------|------------|----------------|-------------|--------|
| Studenti          | 10 (18.18%)    | 12 (21.82%)  | 8 (14.55%) | 10 (18.18%)    | 15 (27.27%) | 55     |
| Jezični model HRV | 12 (60%)       | 8 (40%)      | 0          | 0              | 0           | 20     |
| Ukupno            | 22 (29.33%)    | 20 (26.67%)  | 8 (10.67%) | 10 (13.33%)    | 15 (20%)    | 75     |

Tablica 3: Usporedba ocjena studenata i jezičnog modela (HRV)

Postavljena je null-hipoteza:

$H_0 = N_e$  postoji statistički značajna razlika između distribucija ocjena koje je nakon evaluacije dobio jezični model kada mu je projekt zadan na hrvatskom jeziku u odnosu na ocjene koje su dobili studenti.

Prema  $\chi^2$  test specifikaciji, broj stupnjeva slobode je 4. Granična vrijednost  $\chi^2$  testa sa 4 stupnjeva slobode i razinom signifikantnosti  $p < 0.05$  iznosi 9.49. Izračunata  $\chi^2$  vrijednost za naše podatke iznosi 16.86, što znači da se hipoteza  $H_0$  opovrgava, odnosno postoji statistički značajna razlika između dobivenih ocjena studenata i jezičnog modela.

Slične podatke dobivamo i analizom ocjena kad je projekt bio zadan na engleskom jeziku. Korištenjem  $\chi^2$  testa htjeli smo metodološki potvrditi da postoji i statistički značajna razlika između dobivenih ocjena studenata i jezičnog modela. Distribucije ocjena za studente i jezični model, kad mu je projekt zadan na engleskom jeziku,

prikazan je u tablici 4.

| Ocjena            | Nedovoljan (1) | Dovoljan (2)   | Dobar (3)     | Vrlo dobar (4) | Odličan (5)    | Ukupno |
|-------------------|----------------|----------------|---------------|----------------|----------------|--------|
| Studenti          | 10<br>(18.18%) | 12<br>(21.82%) | 8<br>(14.55%) | 10<br>(18.18%) | 15<br>(27.27%) | 55     |
| Jezični model ENG | 9 (45%)        | 10 (50%)       | 1 (5%)        | 0              | 0              | 20     |
| Ukupno            | 19<br>(25.33%) | 22<br>(29.33%) | 9<br>(12%)    | 10<br>(13.33%) | 15<br>(20%)    | 75     |

Tablica 4. Usporedba ocjena studenata i jezičnog modela (HRV)

Postavljena je null-hipoteza:  $H_0 = \text{Ne postoji statistički značajna razlika između distribucija ocjena koje je nakon evaluacije dobio jezični model kojemu je projekt zadan na engleskom jeziku u odnosu na ocjene koje su dobili studenti.}$

Prema  $\chi^2$  test specifikaciji, broj stupnjeva slobode je 4. Granična vrijednost  $\chi^2$  testa sa 4 stupnjeva slobode i razinom signifikantnosti  $p < 0.05$  iznosi 9.49. Izračunata  $\chi^2$  vrijednost za naše podatke iznosi 13.94, što znači da se hipoteza  $H_0$  opovrgava, odnosno da postoji statistički značajna razlika između dobivenih ocjena studenata i jezičnog modela.

### 3.2 Postoji li razlika u rezultatima koje postiže jezični model na OpenAI platformi kad mu se zada projektni zadatak iz kolegija Java web programiranje u odnosu na prirodni jezik na kojem mu je zadan projektni zadatak (hrvatski ili engleski)?

Na osnovu podataka iz tablice 2, možemo vidjeti kako je prosječna ocjena koju je dobio jezični model kada mu je projekt bio zadan na hrvatskom jeziku bila 1.4, dok je za projekt zadan na engleskom jeziku prosječna ocjena bila 1.6.

Kako nas je zanimalo postoji li statistički značajna razlika između ocjena dobivenih za projekte zadane na oba jezika, izračunali smo  $\chi^2$  test za opažene rezultate iz tablice 2. Obzirom na malene frekvencije ocjena dobar, vrlo dobar i odličan, te kategorije su u svrhu izračuna  $\chi^2$  testa spojene.

Postavljena je null-hipoteza:  $H_0 = \text{Ne postoji statistički značajna razlika između distribucija ocjena koje je nakon evaluacije dobio jezični model kojemu je projekt zadan na engleskom jeziku u odnosu na projekt zadan na hrvatskom.}$

Prema  $\chi^2$  test specifikaciji, broj stupnjeva slobode je tada 2. Granična vrijednost  $\chi^2$  testa sa 2 stupnja slobode i razinom signifikantnosti  $p < 0.05$  iznosi 5.99. Izračunata  $\chi^2$  vrijednost za naše podatke iznosi 0.25, što znači da se hipoteza  $H_0$  potvrđuje, odnosno da ne postoji statistički značajna razlika između dobivenih ocjena jezičnog modela kada je projekt bio zadan na hrvatskom ili na engleskom jeziku.

## 4 Ograničenja

OpenAI jezični modeli istrenirani su na podacima do 2021. godine, uključujući i tu godinu. Iako je istraživanje provedeno na projektnom zadatku koji je generiran 2022. godine i koji je autorsko djelo nastavnika i autora ovoga istraživanja, svejedno postoji mogućnost postojanja sličnih programskih rješenja koja su ušla u podatkovni skup nad kojim je izvršeno treniranje modela.

## 5 Zaključak

U ovom se istraživanju, autor rada fokusirao na metodološke aspekte nastave; konkretno iz područja računalnog programiranja. Trendovi koji se mogu zaključiti iz istraživanja zadnjih par godina su takvi da se nastavni proces sve više transformira iz klasičnog „ex-cathedra“ pristupa ka modernijim pristupima. Neki od primjera su primjena igrifikacije u nastavnom procesu, gdje se iskorištavaju elementi igara kako bise povećao angažman studenata, zatim primjena principa obrnute učionice (engl. flipped classroom) i slično. Osnovna ideja novijih pristupa je studenta transformirati iz pasivnog slušatelja u aktivnog sudionika nastavnog procesa.

Paralelno sa nastavnim procesom, zadnjih godina se intenzivnije mijenja i paradigma vrednovanja stečenog znanja studenata. Specifičan problem u domeni računalnog programiranja je vrednovanje znanja studenata. Tradicionalan pristup u kojemu studenti imaju relativno kratak vremenski period za demonstrirati stečeno znanje često nije dovoljan da bi se kvalitetno vrednovalo njihovo znanje. Istraživanja pokazuju da transformacija procesa vrednovanja znanja na projektni pristup, gdje studenti izrađuju programsko rješenje prema zadanoj specifikaciji, uz dodatnu usmenu verifikaciju, pokazuje pozitivne pomake. Izazov u tom kontekstu može predstavljati pojava jeftinih ili čak besplatnih jezičnih modela baziranih na umjetnoj inteligenciji koji nesavjesnim studentima omogućuju pribavljanje projektnih rješenja koja nisu samostalno izradili.

Autore rada zanimalo je u ovom istraživanju kakve rezultate može postići jezični model, ako mu se zada projektni zadatak na odabranom kolegiju iz domene računalnog programiranja (Web programiranje u Javi), koji dobivaju studenti kao projektni zadatak. Odabran je jezični model daVinci-003 na platformi OpenAI. Istraživanje je pokazalo da jezični model trenutno postižu znatno lošije rezultate u usporedbi s promatranom skupinom studenata koji su rješavali zadani projekt, ali ipak imaju prolaznost na ispitu oko 50%. Stoga bi svaki drugi student mogao iskoristiti jezični model kako bi bez ikakvog samostalnog rada pribavio rješenje projektnog zadatka koje za dovoljava

minimalne kriterije prolaska na kolegiju, a takvo rješenje bi mogao pribaviti bilo kada i to u jako kratkomevremenu.

Ocjenjivanjem ponuđenih rješenja pokazalo se da postoji nekoliko razloga zbog kojih su postojale razlike u bodovanju dobivenih rješenja. Kako je projekt na kojem je napravljeno istraživanje projekt završne godine preddiplomskog studija, neki podrazumijevani zahtjevi koji se uvijek traže od studenata nisu bili dio samog teksta projektnog zadatka. Konkretno, uvijek se od studenata traži da maksimalno poštuju SOLID principe kod razvoja bilo kakvog rješenja. Također, ako se radi o razvoju web sučelja, onda se očekuje da se i HTML, CSS i Javascript komponente oblikuju prema dobrim praksama, izdvajanjem ponavljajućeg kôda u izdvojene datoteke. Nepoštivanje takvih praksi penalizira studentske bodove u nekoj mjeri. U 20 iteracija, jezični model je u nekoliko navrata izgubio određeni postotak bodova zbog rješenja koje nije poštivalo podrazumijevana pravila. Drugi razlog zbog kojeg je dolazilo do razlike u bodovanjima i potencijalnog prolaska ili pada nekog rješenja bio je u nekoliko slučajeva uvjet koji kaže da za prolazak svi ishodi učenja moraju i imati uspješnost veću ili jednaku 50%. U ishodima učenja koje nose manji broj bodova (LO3, LO4, LO5 i LO6), dogodilo da neka od dobivenih rješenja zbog jednog ili pola boda razlike padne neki ishodučenja.

Autore je također zanimalo i ma li odabir jezika na kojem je zadan projektni zadatak utjecaj na rezultat. Jezičnom modelu zadan je projektni zadatak na hrvatskome i engleskome jeziku. Istraživanje je pokazalo da jezični model postiže nešto bolje rezultate na engleskom jeziku, što vjerojatno ima uporište u tome da je veći na programskih rješenja nad kojima je treniran jezični model na engleskome jeziku, ali ne postoji statistički značajna razlika u odnosu na hrvatski jezik.

Prepisivanje je među studenta bilo često i prije pojave jezičnih modela, ali uz današnju dostupnost i mogućnosti koje postojeći jezični modeli imaju, daje nesavjesnim studentima priliku za pristup rješenjima za dataka bez procesa učenja i usvajanje znanja.

Preporuka nastavnicima je stavljanje većeg naglaska na opsežnu verifikaciju studentskih rješenja kako bi se demotiviralo prepisivanje. Jezični modeli kao asistenti u raznim radnim zadacima su poželjni, ali kao pomoć u radu, savladavanju gradiva i procesu učenja; nikako kao mehanizam pribavljanja rješenja bez procesa učenja i usvajanja gradiva. Bez verifikacije rješenja u kontroliranim uvjetima, postoji velika mogućnost da iza nekih predana rješenja studentskih projekata ne stoji nikakav samostalan rad i usvojeno znanje i vještine, što je glavni cilj takvih za dataka.

## Literatura

- Bele, D., Kučak, D., Pašić, Đ. (2023) Increasing Students' Motivation and Improving Outcome by Changing the Course Examination from Test to Project on a University Course. U Skala, K. (ur.) *2023 46th International Convention on Information, Communication and Electronic Technology (MIPRO)* (str. 611-615). IEEE.
- Biswas, S. (2023). ChatGPT and the future of medical writing. *Radiology*, 307(2), e223312.
- Bognar, L. (2012). Kreativnost u nastavi. *Napredak: Časopis za interdisciplinarna istraživanja u odgoju i obrazovanju*, 153(1), 9-20.
- Cigan, V., Šlogar, H. (2012). Istraživanje stavova studenata o metodama poučavanja na visokim čilištima u cilju razvoja poduzetničke kompetencije. *Učenje za poduzetništvo*, 2(1), 179-190.
- Debeljak De Martini, A., Ristić, R., Potočnjak, J. (2020). Nastavne metode u edukaciji zdravstvenih djelatnika. *Journal of Applied Health Sciences= Časopis za primijenjene zdravstvene znanosti*, 6(2), 297-304.
- Dubovicki, S. (2019). Professors' views on the relationship between the curriculum of the teacher education and the development of students' creativity based on the Delphi method—longitudinal research. U Dziurzyński, K.; Duda, E. (ur.), *What is new in the Field of Education?* (str. 61-81). (ISSN 978-83-66010-31-4)
- Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., Prather, J. (2022). The robots are coming: Exploring the implications of openai codex on introductory programming. U *ACE '22: Proceedings of the 24th Australasian Computing Education Conference* (str. 10-19). doi: <https://doi.org/10.1145/3511861.3511863>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312.
- Gratchev, I., Jeng, D. S. (2018). Introducing a project-based assignment in a traditionally taught engineering course. *European Journal of Engineering Education*, 43(5), 788-799.
- Kerić, E., Radanović, I., Lukša, Ž., Garašić, D., Sertić Perić, M. (2017). The effect of active teaching in

- the primary school's ecological content. *Educatio biologiae: časopis edukacije biologije*, 3, 15-20.
- Konecki, M., Lovrenčić, S., Kaniški, M. (2016). Using real projects as motivators in programming education. U *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (str. 883-886). IEEE.
- Kučak, D., Bele, D., Pašić, Đ. (2021). Climbing up the leaderboard: an empirical study of improving student outcome by applying Gamification principles to an object-oriented programming course on a university level. U *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* (str. 527-531). IEEE.
- Kučak, D., Pašić, Đ., Mršić, L. (2022, October). An Empirical Study of Long-Term Effects of Applying Gamification Principles to an Introductory Programming Courses on a University Level. U *Intelligent Computing & Optimization: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)* (str. 469-477). Cham: Springer International Publishing.
- Ljubin-Golub, T., Rijavec, M., Olčar, D. (2020). Student flow and burnout: The role of teacher autonomy support and student autonomous motivation. *Psychological Studies*, 65(2), 145-156.
- Mijwil, M. M., Hiran, K. K., Doshi, R., Dadhich, M., Al-Mistarehi, A. H., Bala, I. (2023). ChatGPT and the future of academic integrity in the artificial intelligence era: a new frontier. *AI-Salam Journal for Engineering and Technology*, 2(2), 116-127.
- Močinić, S. N. (2012). Active teaching strategies in higher education. *Metodički obzori: časopis za odgojno-obrazovnu teoriju i praksu*, 7(15), 97-105.
- Mohorovičić, S., Strčić, V. (2011). An overview of computer programming teaching methods. U *Central European Conference on Information and Intelligent Systems* (str. 47-52). Varaždin: Faculty of Organization and Informatics Varaždin.
- OpenAI Platform (2023). Preuzeto s <https://platform.openai.com/>
- Pap, M., Vdović, R., Baletić, B. (2019). Design Thinking metoda u znanstvenom istraživanju, edukaciji i poslovnoj praksi. *Prostor*, 27 (2(58)), 334-347. doi: [https://doi.org/10.31522/p.27.2\(58\).12](https://doi.org/10.31522/p.27.2(58).12)
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), 07.
- Sullivan, M., Kelly, A., McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 6(1).
- Vinduška, I. (2016). Inovativne edukacijske metode i pristupi u nastavi stručnih predmeta u srednjim medicinskim školama (Doktorska disertacija, University of Zagreb. School of Medicine. Chair of Medical Sociology and Health Economics).
- Yuliansyah, A., Ayu, M. (2021). The implementation of project-based assignment in online learning during covid-19. *Journal of English Language Teaching and Learning*, 2(1), 32-38.
- Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.
- Shukor, N. A., Tasir, Z., Van der Meijden, H., Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844-4853. doi: <https://doi.org/10.1016/j.sbspro.2014.01.1036>
- Steingartner, W., Novitzká, V. (2015). A new approach to operational semantics by categories. U T. Hunjak, V. Kirinić, M. Konecki (ur.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (str. 247-254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

## **Challenges in using the language models of the OpenAI platform in the context of evaluating student solutions to project tasks from the field of computer programming**

### **Abstract**

*The authors of this study have noted for years that students' motivation for continuous work is relatively low. To increase student motivation to practise, University Coledge Algebra faculty have decided to change their approach in recent years. Instead of forcing students to study for an exam on specific dates, they were given projects (assignments) to complete. The research results showed that the results obtained later improved compared to traditional exams. However, with the advent of powerful and free or very low-cost language models such as those of the OpenAI platform, teachers are faced with the challenge that*



students may use such models to solve a project task, which is against the code of ethics. The aim of this research is to determine what kind of results the OpenAI Platform language model can achieve when it is assigned a project task from a specific course. 2 research questions were defined. What result does the language model on the OpenAI platform achieve when it is assigned a project task from the Java Web Programming course and is there a difference in the results compared to the results achieved by the students (Q1)? Is there a difference in the results of the language model compared to the natural language in

which it receives the project task (Croatian or English) (Q2)? The analysis of the results has shown that in certain cases the language model can solve the test without additional user intervention in order to obtain a good grade, although the results obtained are significantly weaker than those of the students. The language model achieves better results when the exam is given in English than in Croatian.

**Keywords:** evaluation of knowledge; project-based learning; OpenAI; academic integrity; Artificial Intelligence.