UMJETNA INTELIGENCIJA I AKADEMSKI INTEGRITET I ČESTITOST: ETIČKE IMPLIKACIJE INTEGRACIJE GENERATIVNE UMJETNE INTELIGENCIJE U SUSTAV VISOKOG OBRAZOVANJA I ZNANOSTI

ARTIFICIAL INTELLIGENCE AND ACADEMIC INTEGRITY AND HONESTY: ETHICAL IMPLICATIONS OF INTEGRATING GENERATIVE ARTIFICIAL INTELLIGENCE INTO THE SYSTEM OF HIGHER EDUCATION AND SCIENCE

Miljana Mićunović Odsjek za informacijske znanosti
Sveučilište Josipa Jurja Strossmayera u Osijeku
mmicunov@ffos.hr
Boris Bosančić Odsjek za informacijske znanosti
Sveučilište Josipa Jurja Strossmayera u Osijeku
bbosancic@ffos.hr

UDK / UDC: 004.8:[174:378]
Izvorni znanstveni rad / Original scientific paper
https://doi.org/10.30754/vbh.68.3.1578
Primljeno / Received 6. 8. 2025.
Prihvaćeno / Accepted 15. 10. 2025.

Sažetak

Cilj: Rad analizira utjecaj sve češće primjene umjetne inteligencije na akademski integritet i čestitost, s posebnim naglaskom na generativnu umjetnu inteligenciju kao čimbenik brojnih etičkih dilema, izazova i rizika. Cilj je rada utvrditi prilike i izazove koje umjetna inteligencija predstavlja za procese učenja, poučavanja i znanstveno-istraživačkog rada u kontekstu akademskog integriteta i čestitosti te ponuditi cjeloviti okvir i smjernice za njezinu odgovornu primjenu.

Pristup/metodologija/dizajn: Korištena je *desk*-metoda pregleda i sinteze relevantne literature, pri čemu su analizirani relevantni stručni i znanstveni radovi te aktualni etički i pravni okviri.

Rezultati: Na temelju toga predloženi su konceptualni okvir i mjere za integraciju umjetne inteligencije u visoko obrazovanje koji, između ostalog, obuhvaćaju pomno planiranje i sustavnu integraciju umjetne inteligencije u proces učenja poučavanja i znanstveno-istraživačkog rada, kontinuirano praćenje i evaluaciju njezina utjecaja na navedene procese, razvoj jasnih institucionalnih politika, kontinuiranu edukaciju studenata i nastavnika te implementaciju alata i obrazovnih politika koji potiču etičnost, odgovornost, transparentnost i akademski integritet i čestitost. U radu se zastupa teza da je umjetnu inteligenciju potrebno promatrati kao asistenta koji podržava ljudski intelektualni, kreativni i originalni znanstveno-istraživački rad, a ne ju izbjegavati ili koristiti kao zamjenu za njih. S obzirom na kontinuirani i gotovo eksponencijalni razvoj tehnologije umjetne inteligencije, posebno se naglašava nužnost stalne procjene etičkih izazova i rizika, osobito u pogledu plagiranja, manipulacije istraživačkim podacima i rezultatima te narušavanja povjerenja u akademsku zajednicu. Unatoč izazovima, uz kreiranje jasnih okvira i smjernica, agilnu obrazovnu politiku, praćenje i kritičku evaluaciju tehnoloških trendova te stalno praćenje i razvoj kulture akademskog integriteta i čestitosti, odgovorna primjena umjetne inteligencije može unaprijediti kvalitetu visokog obrazovanja i znanosti.

Originalnost/vrijednost: Originalnost rada ogleda se u praktičnim implikacijama, odnosno prijedlogu konceptualnog okvira i mjera za visokoškolske ustanove koji pružaju jasne smjernice kako etički i odgovorno integrirati umjetnu inteligenciju u vlastiti rad.

Ključne riječi: akademski integritet i čestitost; etičke implikacije; generativna umjetna inteligencija; konceptualni okvir; visoko obrazovanje

Abstract

Purpose: This paper analyzes the impact of the increasing use of artificial intelligence on academic integrity and honesty, with a special focus on Generative Artificial Intelligence as a source of numerous ethical dilemmas, challenges, and risks. The aim is to identify the opportunities and challenges that artificial intelligence presents for learning, teaching, and research processes in the context of academic integrity and honesty, and to offer a comprehensive framework and guidelines for its responsible use.

Approach/Methodology/Design: A desk-based method of systematic review and synthesis of relevant literature was used, analyzing pertinent professional and scientific works as well as current ethical and legal frameworks.

Findings: Based on the analysis, the paper proposes a conceptual framework and measures for integrating artificial intelligence into higher education. These include careful planning and systematic integration of artificial intelligence into the processes of learning, teaching, and research; continuous monitoring and evaluation of its impact; development of clear institutional policies; continuous education of students and teaching staff; and the implementation of tools and educational policies that promote ethical behaviour, responsibility, transparency, and academic integrity and honesty. The paper argues that artificial intelligence should be viewed as an assistant that supports human intellectual, creative, and original research work, rather than something to be avoided

or used as a replacement. Given the continuous and nearly exponential development of AI technology, the necessity of ongoing assessment of ethical challenges and risks is particularly emphasized—especially regarding plagiarism, manipulation of research data and results, and the erosion of trust within the academic community. Despite these challenges, the responsible use of artificial intelligence supported by the development of clear frameworks and guidelines, agile educational policy, monitoring and critical evaluation of technological trends, and the continuous development of a culture of academic integrity and honesty, can enhance the quality of higher education and science.

Originality/Value: The originality of this paper lies in its practical implications, namely in the proposed conceptual framework and measures for higher education institutions that provide clear guidelines on how to ethically and responsibly integrate artificial intelligence into their work.

Keywords: academic integrity and honesty; conceptual framework; ethical implications; generative artificial intelligence; higher education

1. Umjesto uvoda – Osvrt na povijest umjetne inteligencije iz perspektive zaštite akademskog integriteta i čestitosti

U članku *As We May Think*, objavljenom davne 1945. godine, Vannevar Bush ponudio je rješenje za tadašnji problem informacijske eksplozije tiskanih publikacija, koja je dosezala svoj vrhunac. Rješenje je vidio u razvoju informacijske tehnologije. Zamislio je stroj koji je nazvao *Memex*, i koji je, prema njegovoj zamisli, trebao zaprimiti ulogu asistenta istraživača ili znanstvenika, pomažući mu u radu (Bush, 1945). U kontekstu ovoga rada izvorni koncept *Memexa* važan je jer ocrtava prvotni smjer razmišljanja u kojem se svaki oblik informacijske tehnologije – pa i onaj koji će se tek pojaviti – promatra kao asistent, pomagač, a ne kao netko ravnopravan ili nadređen čovjeku u znanstvenim istraživanjima.

Međutim, uskoro su se pojavili drugačiji pogledi na informacijsku tehnologiju koji su otvorili prostor drugačijem smjeru razmišljanja o njezinoj ulozi u znanstvenom istraživanju. Premda se razlike između biološke i strojne inteligencije u to vrijeme nisu ozbiljnije razmatrale, čemu svjedoči i rad McCullocha i Pittsa (1943) o konceptu „umjetnog neurona“, koji je među biolozima naišao na uglavnom „hladan prijem“, njihov je rad ipak privukao pozornost računalnih inženjera. Među tim novim pogledima na informacijsku tehnologiju često se spominje i pogled Norberta Wienera koji je u istoimenoj knjizi iznio tezu da ljudski mozak možda funkcionira poput računala (Wiener, 1948).

Ipak prve konkretne teorijske postavke vezane za funkcioniranje računala kao inteligentnog entiteta postavio je britanski znanstvenik Alan Turing u radu *Computing machinery and intelligence*, objavljenom 1950. godine (Turing, 1950). Možda su ga na to potaknuli i razgovori s utemeljiteljem, po mnogima, „teorije informacije“ Claudeom E. Shannonom tijekom Drugoga svjetskog rata, o pitanju

mogu li računala misliti, što spominje Gleick (2011) u svojoj knjizi o pojmu informacije. Vrijedi spomenuti i da je Turing smatrao mozak računalom koje funkcionira zahvaljujući procesuiranju informacija (Adams, 2003: 474).

U tom pionirskom radu o tzv. „računalnoj inteligenciji", Turing je predložio test koji bi mogao pokazati može li stroj odgovarati tako da ga ispitivač ne može razlikovati od čovjeka? U izvorniku, test je nosio naziv *igra oponašanja* (engl. *imitation game*), no kasnije je postao poznat kao Turingov test. U testu ljudski ispitivač preko terminala ili pisaćeg stroja komunicira s računalom i drugim čovjekom. U trenutku kada više nije siguran pripada li odgovor računalu ili čovjeku – računalo je prošlo test (Turing, 1950).

Za razliku od Busha, koji je informacijsku tehnologiju promatrao tek kao pomoćni alat, Turing ju je na temelju njezina najingenioznijeg oblika – umjetne inteligencije – nastojao uzdići na čovjekovu razinu, ne sluteći kakve će posljedice ta odluka imati u budućnosti, osobito kad je riječ o akademskom pisanju, a posljedično i akademskom integritetu. Razvidno je da su istraživači prirodnih i tehničkih znanosti od samog početka skloni u umjetnoj inteligenciji vidjeti nesagledIv potencijal koji će ju jednog dana ne samo izjednačiti s ljudskom inteligencijom (proći Turingov test) već i nadmašiti. Suprotno tome stručnjaci iz društvenih i humanističkih znanosti taj „tehnološki optimizam" nikada nisu dijelili.

Britanski filozof John Searle tu je podjelu dodatno artikulirao razgraničenjem umjetne inteligencije na „jaku" (engl. *strong*) i „slabu" (engl. *weak*) (Searle, 1980). Pristalice „jake" umjetne inteligencije vjerovali su da će ona ne samo dostići ljudsku inteligenciju već je i nadići, dok pobornici takozvane „slabe" umjetne inteligencije u njoj nisu vidjeli ništa što bi moglo ozbiljno konkurirati ljudskim sposobnostima. No svojim misaonim eksperimentom Kineske sobe, Searl je skrenuo pažnju na još nešto, puno važnije: računala**,** pokazao je Searl, funkcioniraju tako da ne moraju ništa razumjeti. Zamislio je osobu engleskog govornog područja (sebe) zatvorenu u sobi i kojoj je dana „velika hrpa kineskog pisma" koje ne razumije. Vrlo pojednostavljeno, zadatak te osobe bio je da na temelju dobivenih kineskih znakova i slijedeći pravila iz priručnika na engleskom jeziku koji joj je dostupan, odašilje druge kineske znakove kao odgovor. Eksperiment je pokazao da, usprkos tomu što osoba zatvorena u Kineskoj sobi znakove koje dobiva ne razumije, ipak uspješno izvršava zadatak koji je joj je zadan (Searl, 1980).

Ako je Searle u pravu, onda je svojim argumentom srušio nade Pittsa, McCullocha, Wienera i Turinga, koji su na ljudski mozak gledali kao na računalo.

Bilo kako bilo, pokazalo se da na Searlovim postavkama funkcionira najmanje jedan oblik umjetne inteligencije koji se u međuvremenu pojavio – nakon niza godina razvoja oblika temeljenih na pravilima za manipulaciju simbolima (ekspertni sustavi su najeklatantniji primjer). Premda ne posve na način kako ga je

opisao John Searle, ali svakako bez ikakvog uključenog razumijevanja, razvojem dubokih neuronskih mreža – osobito od 2011. godine – počeo se oblikovati nov, poseban oblik umjetne inteligencije koji će se ubrzo pokazati revolucionarnim: generativna umjetna inteligencija (engl. *Generative Artificial Intelligence* – GAI). Za razliku od dotadašnjih simboličkih pristupa (kakvi su bili ekspertni sustavi), na koje se pozivao i Searle, generativna umjetna inteligencija se oslanja na nešto što se na prvi pogled čini paradoksalno jednostavnim: predviđanje sljedeće riječi u nizu na temelju statističkih obrazaca u jeziku. Ta, naizgled jednostavna metoda, ima i svoju (pret)povijest koju ovdje vrijedi istaknuti.

Već u 19. stoljeću, Samuel F. B. Morse i Alfred Veil uočili su da im statistička učestalost slova engleske abecede može poslužiti u izgradnji učinkovitije Morseove abecede, odnosno kôda. Najčešćim slovima dodijelili su najkraće sekvence točaka i crtica, čime su znatno skratili vrijeme potrebno za slanje poruka telegrafom (Bosančić, 2023: 56–57; Gleick, 2011). Shvaćajući da statističke zakonitosti jezika omogućuju veću predvidivost, a time i učinkovitiji prijenos informacija, Claude E. Shannon ih je 1948. godine iskoristio kao temelj u svojoj znamenitoj *Matematičkoj teoriji komunikacije* (u kojoj su mnogi prepoznali utemeljujuću teoriju pojma informacije). Kako bi pokazao da se predvidivost jezika može koristiti za slanje poruka, Shannon je proveo statističku analizu učestalosti pojavljivanja pojedinačnih slova, parova i trojki slova, kao i jedne ili dviju riječi u engleskim rečenicama. Na temelju tih rezultata razvio je postupak predviđanja pojave simbola ovisno o prethodnima, koji je nazvao aproksimacijama (Bosančić, 2023).

Upravo operacija predviđanja pojave sljedeće riječi ili simbola u tekstu (kao poruci), ovisno o riječi ili simbolu koji joj prethodi, predstavlja temelj na kojem počiva i današnja generativna umjetna inteligencija. No kontekst razvoja generativne umjetne inteligencije nepotpun je bez spomena dubokog učenja, sofisticiranijeg oblika strojnog učenja te dubokih neuronskih mreža, čiji su potencijal razvili i popularizirali Geoffrey Hinton i suradnici, pri čemu je Hinton za taj doprinos nagrađen i Nobelovom nagradom 2024. godine. U svom ključnom radu *Learning representations by back-propagating errors* iz 1986. godine, opisali su temeljni algoritam za učenje u takvim višeslojnim ili dubokim neuronskim mrežama, poznat kao *backpropagation*, postavivši temelje za treniranje jezičnih modela koji će se pojaviti u 21. stoljeću (Rumelhart, Hinton i Williams, 1986).

Počevši od 2010-ih, s napretkom računalne snage i dostupnošću hardvera za brzu obradu podataka, GAI je, u obliku velikih jezičnih modela (engl. *Large Language Models* – LLMs), započeo generirati tekstove slične onima koje bi napisao čovjek, i to upravo predviđanjem sljedeće riječi u nizu u odnosu na prethodnu riječ, i u odnosu na druge riječi u širem kontekstu. Ti su modeli trenirani na sveobuhvatnom korpusu tekstova koji obuhvaćaju različita područja ljudskog znanja, a njihova učinkovitost proizlazi iz milijardi parametara koji su fino podešeni kroz

iterativni proces treniranja, što ih, u konačnici, i usmjerava prema sve relevantnijim i uvjerljivijim odgovorima. Rezultat je bio i šokantan i uznemirujući: generativna umjetna inteligencija je, bez ikakvog razumijevanja, uspjela generirati tekstove koji su po stilu i sadržaju bili gotovo nerazlučivi od onih koje bi napisao čovjek – i to ne bilo tko, već ekspert u tom području.

Ali, je li to uistinu bio znak da je generativna umjetna inteligencija prošla Turingov test? Iako su mnogi vjerovali da je s pojavom ChatGPT-a 2022. godine generativna umjetna inteligencija konačno zadovoljila taj kriterij, to se nije dogodilo. Premda briljira u generiranju teksta, taj jezični model još uvijek nije sposoban dosljedno i pouzdano izvoditi zaključke. Zbog svoje parcijalne uspješnosti u oponašanju ljudske inteligencije i rješavanju kompleksnih zadataka, razvidno je da Turingov test više ne predstavlja pouzdan kriterij za procjenu inteligencije tih sustava, i da je potrebno osmisliti nove oblike evaluacije (Biever, 2023).

Bilo kako bilo, s pojavom generativne umjetne inteligencije javila se i potreba za preformuliranjem izvorne Searleove podjele na „jaku" i „slabu" umjetnu inteligenciju. Budući da generativna umjetna inteligencija još uvijek nije dosegnula razinu kognitivnih sposobnosti usporedivu s ljudskom inteligencijom, smatra se da bi to trebala učiniti opća umjetna inteligencija (engl. *Artificial General Intelligence* – AGI), koja bi u nekim aspektima ljudsku inteligenciju mogla čak i prevladati (Bubeck i sur., 2023). Iako se u javnom prostoru sve češće iznose tvrdnje kako bi opća umjetna inteligencija mogla postati stvarnost već u narednim godinama (Marcus, 2023; Yudkowsky, 2023), među znanstvenicima još uvijek ne postoji konsenzus. Pritom predviđanja dosežu od sljedećeg desetljeća pa sve do kraja stoljeća (Grace i sur., 2018). Zanimljivo je da se ni tu ne zaustavljaju projekcije o razvoju umjetne inteligencije: nakon opće umjetne inteligencije, sljedeća (zasad u potpunosti hipotetska) faza pripisuje se superumjetnoj inteligenciji (engl. *Artificial Super Intelligence* – ASI), za koju se predviđa da bi mogla u potpunosti nadmašiti ljudsku inteligenciju u svim područjima, uključujući znanstvenu kreativnost, emocionalnu inteligenciju i druge aspekte kognitivnog funkcioniranja (Bostrom, 2014).

Namjera prikaza povijesne analize razvoja umjetne inteligencije u kontekstu zaštite akademskog integriteta i čestitosti bila je odgovoriti na pitanje „Kako smo došli do toga da cijenu veće učinkovitosti u znanstvenoj produkciji plaćamo dovođenjem u opasnost akademski integritet i čestitost?" te ukazati na jednu važnu činjenicu: da generativna umjetna inteligencija, u svojem današnjem obliku, još uvijek nije u potpunosti prošla Turingov test; štoviše, da je u brojnim aspektima simulacije ljudske inteligencije pokazala ozbiljne manjkavosti (u apstraktnom mišljenju, logičkom zaključivanju, simulaciji razumijevanja). No istodobno u području akademskog pisanja generativna umjetna inteligencija u velikoj je mjeri i

neočekivano pokazala visoku razinu uspješnosti, čime je i otvorila pitanje zaštite akademskog integriteta i čestitosti.

Hoćemo li u budućnosti krenuti putem poboljšanja vlastitih kognitivnih sposobnosti (transhumanistički pristup) i dopustiti integraciju umjetne inteligencije s biološkom inteligencijom – o tome zasad ne možemo ništa znati. No sve dok se to ne dogodi, možemo nastaviti putem koji smo već započeli: svaki oblik umjetne inteligencije koji se pojavi nastojati prije integrirati nego zabraniti. To se osobito odnosi na područje akademskog pisanja, u kojem je nužno što prije iznaći način korištenja umjetne inteligencije koji bi se temeljio isključivo na podršci i asistenciji, u skladu s načelima akademskog integriteta i čestitosti, možda i posebno prilagođenima za tu novu zadaću.

2. Metodologija

U radu je primijenjena *desk*-metoda temeljena na pregledu relevantne stručne i znanstvene literature te aktualnih etičkih i pravnih okvira i akata koji su se odnosili na pitanje primjene (generativne) umjetne inteligencije i velikih jezičnih modela u visokom obrazovanju, s posebnim naglaskom na etičke implikacije i njihov utjecaj na akademski integritet i čestitost. Pregled i odabir literature proveden je kombinacijom pretrage baza podataka Scopus i Web of Science te uvida u stručne i znanstvene publikacije iz osobne arhive autora. Dodatno su analizirani pravni i strateški dokumenti objavljeni na mrežnim stranicama institucija Europske unije, posebice Europske komisije. Također su analizirani i odabrani mrežni izvori, priručnici i izvještaji relevantnih međunarodnih i nacionalnih tijela. Pretraga literature provedena je od svibnja do srpnja 2025. godine. Pretraga literature u bazama podataka provedena je prema zadanim strategijama:

- u bazi podataka Scopus: TITLE-ABS-KEY(„artificial intelligence“ AND „academic integrity“) AND PUBYEAR > 2020 AND (LIMIT TO(DOCTYPE, „ar“) AND LANGUAGE, („English“) te
- u bazi podataka Web of Science: „artificial intelligence“ (Topic, TS) AND „academic integrity“ (Topic, TS), Articles (Document Types, DT), English (Languages), 2020-01-01 to 2025-05-01 (Publication Date), Relevance (Sort By).

Pretraga je nadopunjena metodom „snježne grude“ (engl. *snowballing, citation chainin*g), odnosno utvrđivanjem dodatnih relevantnih izvora uz pomoć bibliografija pronađenih radova. U analizu su uključeni izvori koji su izravno povezani s temom istraživanja i pružaju teorijski, konceptualni ili empirijski doprinos razumijevanju etike umjetne inteligencije u kontekstu akademskog integriteta i čestitosti. Isključeni su izvori koji se ponavljaju, čiji su sadržaji irelevantni ili zastarjeli te radovi bez provjerljivih referenci. Odabrana su ukupno 92 izvora objavljena od

1948. do 2025. godine: 53 znanstvena članka, 3 članka u zbornicima skupova, 9 monografija, 6 poglavlja u knjigama, 5 izvještaja, 5 pravnih akata (zakoni, smjernice, preporuke, i dr.), 6 priručnika/vodiča, 5 mrežnih izvora.

Kritička analiza literature obuhvaćala je ponajprije pitanje etičkih implikacija općenito, ali i pitanje ugroze akademskog integriteta i čestitosti, od prilagodbe pedagoških praksi i kurikuluma do institucionalne odgovornosti. Tako je omogućen cjeloviti uvid u aktualna saznanja, preporuke i prakse u tom području te su utvrđeni izazovi i prilike za trenutačnu i buduću implementaciju umjetne inteligencije u visokom obrazovanju. U konačnici, ključne ideje, koncepti i zaključci sintetizirani su kako bi se stvorili temelji za daljnji teorijski i praktični rad koji je obuhvaćao dizajn cjelovitog konceptualnog okvira i mjera potrebnih za uspješnu, učinkovitu, etičku i odgovornu integraciju umjetne inteligencije u sustav visokog obrazovanja u Republici Hrvatskoj. Predloženi konceptualni okvir i mjere utemeljeni su na tezi da je umjetnu inteligenciju nužno integrirati kao suvremenu i korisnu tehnologiju i koristiti kao asistenta u učenju, poučavanju i istraživanju, a ne ju izbjegavati ili koristiti kao zamjenu za čovjekov kognitivni, kreativni i originalni znanstveni rad.

3. Uvid u europski i međunarodni zakonodavni okvir u području umjetne inteligencije

Mogućnosti (generativne) umjetne inteligencije utječu na svaki segment društva i oblikuju ga, od gospodarstva, ekonomije i politike, preko obrazovanja, znanosti i kulture, do slobodnog vremena i zabave. Usporedno s brzim napretkom tehnologije generativne umjetne inteligencije raste i potreba za jasnim pravilima koja će regulirati i usmjeravati njezin razvoj i primjenu. U odnosu na to često je pitanje kako uspostaviti ravnotežu između tehnološke inovacije i zaštite njezinih korisnika i društva općenito. Pravna regulacija određene tehnologije svakako nije novo pitanje, no naglim razvojem generativne umjetne inteligencije ponovno je postalo aktualno pitanje uspostave ravnoteže između zaštite korisnika i tržišta s jedne strane (primjerice, postavljanje minimalnih standarda privatnosti, sigurnosti i transparentnosti), poticanja daljnjeg razvoja tehnologije s druge strane (poznato je da prestroge regulacije mogu usporiti inovacije i odvratiti ulaganja u razvoj i usavršavanje tehnologije) i održavanja konkurentnosti među tvrtkama i organizacijama koje se bave razvojem i proizvodnjom tehnologije s treće strane (važno je spriječiti monopolizaciju tržišta i omogućiti i manjim tvrtkama ulazak na tržište). Zbog toga se pravna regulacija često uvodi postupno, putem određenih smjernica, standarda i pilot-studija, prije nego što se uspostavi čvrsti zakon koji će vrijediti za sve.

Kada je u pitanju tehnologija generativne umjetne inteligencije, postoje ključne inicijative i regulacije na području Europske unije i šire. Cilj je takvih

inicijativa osigurati razvoj umjetne inteligencije u skladu s temeljnim pravima i demokratskim vrijednostima, odnosno spriječiti zlouporabu tehnologije i potaknuti povjerenje javnosti. Na području Europske unije, uz poznatu *Bijelu knjigu o umjetnoj inteligenciji*, kojom Europska komisija jamči okvir za pouzdanu umjetnu inteligenciju temeljen na izvrsnosti i povjerenju (Europska komisija, 2020) i *Etičke smjernice za pouzdanu umjetnu inteligenciju* (AI HLEG, 2019) u kojima se iznosi sedam ključnih zahtjeva za pouzdanu umjetnu inteligenciju, trenutačno je na snazi *Akt o umjetnoj inteligenciji* kojim se nastoji osigurati niz mjera za odgovoran razvoj pouzdane umjetne inteligencije. *Aktom* se tehnologija umjetne inteligencije kategorizira prema četiri razine rizika: 1) minimalni rizik (npr. video igre i *e*-pošta koji koriste tehnologiju umjetne inteligencije i koji nisu zahvaćeni *Aktom*), 2) ograničeni rizik (npr. *chatbotovi* i sustavi generativne umjetne inteligencije kojima *Akt* propisuje nužno obavješćivanje korisnika o generiranom sadržaju kako bi korisnik mogao donijeti informiranu odluku o daljnjem korištenju sadržaja), 3) visoki rizik (npr. autonomna vozila, tehnologija za biometrijsku identifikaciju i medicinski uređaji za dijagnostiku koji, prema *Aktu*, moraju ispunjavati stroge zahtjeve i obveze tržišta Europske unije) i 4) neprihvatljivi rizik (npr. sustavi umjetne inteligencije koji ugrožavaju ljudska prava i njihovu egzistenciju i koje je zabranjeno koristiti u zemljama Europske unije). Primjerice, sustavi visokog rizika podliježu strogim zahtjevima upravljanja rizicima, tehničkog dokumentiranja razvoja i primjene tehnologije te usklađenosti s drugim povezanim zakonodavnim okvirima (npr. *Zakon o autorskom pravu*), pa je tako jasno propisano korištenje sustava za biometrijsku identifikaciju koji se ne smiju koristiti u javnim prostorima u stvarnom vremenu, osim u opravdanim slučajevima od javnog interesa (očuvanje javne sigurnosti, potraga za nestalim osobama, i sl.). Uz regulaciju umjetne inteligencije na temelju zakonodavstva o temeljnim ljudskim pravima i sigurnosti, *Akt* nastoji regulirati, odnosno poticati ulaganja i razvoj inovacija u području umjetne inteligencije (European Union, 2024). Uvođenje jedinstvenog europskog pravnog okvira moglo bi ojačati poziciju Europske unije kao globalnog lidera u području etike umjetne inteligencije. To bi, s jedne strane, povećalo povjerenje europskih građana i poduzetnika, ali i postavilo visoke standarde za strane tvrtke koje žele poslovati na europskom tržištu (European Commission, 2021; European Union, 2024). Iako se *Akt* eksplicitno ne bavi pitanjem ugroze akademskog integriteta i čestitosti uslijed integracije generativne umjetne inteligencije u visoko obrazovanje, upravo je druga razina rizika, koja podrazumijeva obavješćivanje o sadržaju generiranom umjetnom inteligencijom, povezana s pitanjem očuvanja akademskog integriteta.

Kina je do 2020. godine uspostavila inicijalne korake u kreiranju etičkih normi, politika i zakonodavstva u području umjetne inteligencije, a do 2030. godine planira imati cjeloviti zakonodavni okvir koji bi uključivao zakone, regulativne norme, etičke norme i politike za odgovorni razvoj umjetne inteligencije. Kineski

model uključuje četiri ključna aspekta okvira za razvoj odgovorne umjetne inteligencije: 1) podatke, 2) algoritme, 3) platforme i 4) primjenu umjetne inteligencije u različitim specifičnim kontekstima (primjerice, korištenje umjetne inteligencije za biometrijsku analizu pri ulasku u stambene zgrade). Slično kao i *Akt o umjetnoj inteligenciji*, različiti tehnički standardi, etička načela, inicijative i smjernice u Kini oslanjaju se na postojeći zakonodavni okvir, tj. na *Zakon o e-trgovini*, *Zakon o zaštiti podataka* i *Zakon o zaštiti osobnih informacija* (Shen i Liu, 2022). U Sjedinjenim Američkim Državama (SAD) određene vladine agencije već igraju ključnu ulogu u poboljšanju automatiziranih sustava i razvoju tzv. humanocentrične umjetne inteligencije, poput Američke agencije za hranu i lijekove (engl. *Food and Drug Administration* – FDA), Savezne uprave za zrakoplovstvo (engl. *Federal Aviation Administration* – FAA), Savezne trgovinske komisije (engl. *Federal Trade Commission* – FTC), i dr. Iako je vladina intervencija u regulaciji tehnologije umjetne inteligencije nužna u industrijama kao što su proizvodnja medicinskih uređaja i lijekova te skrb o bolesnima i starijima, postoji i određeni otpor prema vladinoj regulaciji zbog spomenutih rizika i negativnih utjecaja regulacije (Shneiderman, 2022). Na međunarodnoj razini važno je spomenuti i *Preporuke Organizacije za gospodarsku suradnju i razvoj* (OECD) iz 2019. godine koje predstavljaju prvi međunarodni standard u području politike umjetne inteligencije. Preporuke promiču načela poput odgovornosti, transparentnosti, robusnosti, uključivog razvoja i humanocentrične umjetne inteligencije. Za razliku od *Akta o umjetnoj inteligenciji*, OECD-ove *Smjernice* ne oslanjaju se na postojeće obvezujuće zakonske okvire, već pružaju smjernice za suradnju na nacionalnoj i međunarodnoj razini, posebice u području istraživanja i razvoja, izgradnje ljudskog kapaciteta i promicanja digitalnih ekosustava (OECD, 2019). Pregled postojećih zakonskih rješenja za odgovorni razvoj i primjenu tehnologije umjetne inteligencije u Europi, Kini i SAD-u upućuje na postojanje dvaju glavnih pristupa. Prvi se odnosi na kreiranje zakonskih okvira na temelju najbržeg mogućeg predviđanja razvoja tehnologije umjetne inteligencije. Naime nakon osnovne prvotne primjene tehnologije, zakonodavci mogu sažeti i predvidjeti moguće rizike na temelju dane situacije i u skladu s tim kreirati zakonodavni okvir. Drugi se pristup odnosi na „čekanje" (tzv. „pričekajmo pa ćemo vidjeti" pristup) koji smatra da je trenutačno prerano da zakonodavci razumiju kako će tehnologija umjetne inteligencije zaista utjecati na život i rad građana. S obzirom na to, zakonodavci se fokusiraju više na pozitivni utjecaj tehnologije, dok će se rizici sami detektirati, prilagoditi, regulirati i riješiti kroz mehanizme slobodnog tržišta i konkurentnosti (Shen i Liu, 2022). U kontekstu izazova koje umjetna inteligencija predstavlja za akademski integritet i čestitost i s obzirom na činjenicu da slobodno tržište i načela konkurentnosti neće riješiti probleme akademske zajednice, prvi se pristup čini izglednijim.

Na globalnoj razini moguće je zamijetiti rastući broj dionika koji se uključuje u oblikovanje pravila i politika za razvoj i primjenu umjetne inteligencije. Uspored-

ba Europe, Kine i SAD-a ukazuje na različite pristupe i različite strategije pravne regulacije umjetne inteligencije, što bi moglo otežati međunarodnu suradnju i potaknuti natjecanje među različitim dionicima u postavljanju globalnih standarda. Ono što je svakako sigurno jest da regulacije umjetne inteligencije polako prelazi iz teorije u praksu i da zakonodavne mjere koje se danas oblikuju definiraju budućnost te tehnologije i njezin utjecaj na društvo i gospodarstvo u budućnosti.

Dok zakonodavni i pravni okvir pruža formalne smjernice i ograničenja za razvoj i primjenu umjetne inteligencije, etika umjetne inteligencije otvara šira pitanja odgovornosti i pravednosti koji utječu na pojedinca i društvo. Ta pitanja, između ostalog, posebno dolaze do izražaja u kontekstu (visokog) obrazovanja i znanosti, gdje integracija umjetne inteligencije utječe na rad visokoškolskih ustanova i na procese učenja, poučavanja i znanstveno-istraživačkog rada.

4. Etika umjetne inteligencije u visokom obrazovanju i znanosti

4.1. Općenito o etici i moralnosti u području umjetne inteligencije

Kako se umjetna inteligencija počela šire integrirati u društvo, od svakodnevnih aplikacija do složenih sustava u medicini, prometu i financijama, tako je pitanje pouzdane umjetne inteligencije postalo sve važnijim. Pouzdana umjetna inteligencija omogućuje odgovorno korištenje tehnologije u različitim sektorima i situacijama, a predviđa se i njezina velika uloga u rješavanju društvenih problema, poput klimatskih promjena i rastuće socijalne nejednakosti. No da bi umjetna inteligencija zaista bila korisna, mora se postaviti jasan etički okvir koji će ju uskladiti s društvenim vrijednostima i normama.

Jedan od prijedloga kako postići pouzdanu umjetnu inteligenciju jest tzv. piramida etičke umjetne inteligencije (Schmarzo, 2023). Radi se o proaktivnom etičkom okviru koji stavlja naglasak na tri ključna načela: transparentnost (razumijevanje kako sustavi umjetne inteligencije donose odluke), nepristranost (izbjegavanje diskriminacije i pristranih rezultata) i kontinuirano učenje sustava (stalna prilagodba sustava društvenim promjenama i novim etičkim izazovima).

Važno je ne svoditi etiku umjetne inteligencije samo na tehnička pitanja poput pristranosti algoritama ili objašnjivosti rada određenog sustava i modela umjetne inteligencije. Ona mora obuhvatiti društvene i političke implikacije i pitanja poput toga kako umjetna inteligencija utječe na poslove i radna mjesta, na demokraciju, privatnost i svakodnevni život čovjeka. Upravo zbog toga važan je cjeloviti pristup definiranju i uspostavljanju etike umjetne inteligencije, pristup koji će uključivati tehničke, društvene, političke i filozofske aspekte (Borenstein i Howard, 2021; AI HLEG, 2019).

Uz navedena etička pitanja, a zahvaljujući sve češćim predviđanjima skorašnjeg razvoja opće umjetne inteligencije, pa čak i superinteligencije, postavlja se i pitanje moralnog statusa umjetne inteligencije. Coeckelbergh (2020: 47–62) razmatra to pitanje razlikujući dva pojma, odnosno dvije moralne uloge umjetne inteligencije – one moralnog agenta i moralnog pacijenta. Uloga moralnog agenta pretpostavlja da će umjetna inteligencija moći samostalno moralno djelovati i snositi odgovornost za svoje postupke, dok uloga moralnog pacijenta pretpostavlja da je umjetna inteligencija tzv. trpitelj moralnog djelovanja, tj. ona prema kojoj se moralno djeluje i koja zaslužuje određena moralna prava. Po pitanju moralne agencije postoji više struja. Jedni tvrde da umjetna inteligencija nikada ne može biti moralni agent jer joj nedostaju svijest i emocije, drugi priznaju djelomičnu moralnu sposobnost sustava umjetne inteligencije, dok treći smatraju da napredni sustavi, poput autonomnih vozila, već sada pokazuju funkcionalnu sposobnost moralnog prosuđivanja, tj. u stanju su procijeniti posljedice svojih odluka i djelovanja te se ponašati u skladu s etičkim pravilima koje su im postavili ljudi. U odnosu na umjetnu inteligenciju kao moralnog pacijenta važna je čovjekova emocionalna reakcija i djelovanje spram sustava umjetne inteligencije. To se, naravno, većinom odnosi na robote koji predstavljaju svojevrsno „utjelovljenje“ umjetne inteligencije. Određena istraživanja pokazuju sposobnost ljudi da emocionalno reagiraju prema robotima i strojevima, pa čak i da osjećaju nelagodu kada ih „povrjeđuju“ ili svjedoče njihovom „zlostavljanju“. Već su filozofi 18. stoljeća, poput Immanuela Kanta, upozoravali da čovjekovo loše ponašanje prema neljudskim bićima i entitetima može narušiti njegov moralni karakter i tako smanjiti njegovu sposobnost suosjećanja i prema drugim ljudima. Znanstvenici koji predviđaju razvoj superinteligencije koja će imati razvijenu svijest vjeruju da bi ona mogla steći izravni moralni status. U konačnici, važno je i pitanje kriterija dodjele moralnog statusa koji su se svakako mijenjali kroz povijest. Ljudi su skloni hijerarhijskom vrednovanju entiteta, što može dovesti do patronizirajućeg odnosa prema umjetnoj inteligenciji, pa neki autori smatraju da je važno preispitati ljudske pretpostavke o moralnosti i izbjegavati praksu u kojoj su ljudska iskustva jedini standard i mjerilo etičkog vrednovanja (Coeckelbergh, 2020: 47–62). Ako općenito pitanje etike i moralnosti umjetne inteligencije stavimo u kontekst znanosti i akademskog integriteta i čestitosti, možemo si postaviti pitanje tko bi snosio odgovornost i koji bi etički okvir bio prikladan odgovor na situaciju u kojoj superinteligentni sustav sudjeluje u provođenju eksperimentalnog medicinskog istraživanja analizirajući velike skupove podataka (npr. ispitivanje učinka lijeka na karcinom jednjaka) i lažira podatke i rezultate istraživanja kako bi privukao što više ulaganja ili zadovoljio unaprijed postavljene uvjete projektnog natječaja?

4.2. Umjetna inteligencija u visokom obrazovanju i znanosti

Autori Ouyang i Jiao (2021) u svom radu izdvajaju tri glavne paradigme primjene umjetne inteligencije u obrazovanju:

1. obrazovanje usmjereno umjetnom inteligencijom (tehnologija umjetne inteligencije vodi proces učenja i poučavanja)
2. obrazovanje potpomognuto umjetnom inteligencijom (tehnologija umjetne inteligencije podržava studente i nastavnike u njihovu radu) i
3. obrazovanje osnaženo umjetnom inteligencijom (tehnologija umjetne inteligencije koristi se za jačanje autonomije studenata i nastavnika i personalizaciju učenja, pri čemu studenti postaju aktivni kreatori vlastita obrazovanja).

Većina se današnjih trendova u obrazovanju i znanosti vodi idejom treće paradigme, odnosno idejom korištenja alata umjetne inteligencije za osnaživanje studenata, nastavnika i znanstvenika.

Generativna umjetna inteligencija, posebice *chatbotovi* kao što su ChatGPT, Gemini, Perplexity, Claude, DeepSeek, Grok i drugi, postaju sve popularnijim alatima u visokom obrazovanju i znanosti. Razloga je dosta. Tehnologija generativne umjetne inteligencije podržava uvođenje novih pedagoških pristupa i praksi. Kako pokazuju radovi grupe autora (Arowosegbe, Alqahtani i Oyelade, 2024; Chiu i sur., 2023; Cordona, Rodriguez i Ishmael, 2023; Kralj i sur., 2024; Mrnjaus, Vrcelj i Kušić, 2023; Msambwa, Wen i Daniel, 2025; Teachflow, 2023), (generativna) umjetna inteligencija podržava personalizirano učenje i prilagođava nastavu vještinama i potrebama pojedinog studenta, dodatno motivira i angažira studente, podržava inkluziju smanjujući prepreke u obrazovanju za studente s posebnim potrebama, a ujedno podržava i kreativni i intelektualni rad studenata tako što im pomaže u oblikovanju ideja i pisanju. Ipak, zanimljivo je istraživanje Wangsa i suradnika (2024) koje ukazuje na određene razlike među različitim *chatbotovima*, tj. činjenicu da se *chatbotovi* ne mogu jednako uspješno koristiti u svim znanstvenim područjima – neki su učinkovitiji, a neki slabiji, ovisno u koju se svrhu koriste i u kojem znanstvenom području. Sve se češće zamjećuje i trend kombiniranja umjetne inteligencije s metodama učenja temeljenog na projektima i rješavanju problema te sa zadacima koji zahtijevaju kritičko razmišljanje i kreativni pristup (Chan i Colloton, 2024). Nastavnicima je, pak, najveća podrška pomoć u planiranju nastave i strukturiranju nastavnih sadržaja, olakšavanje administrativnih zadataka, vrednovanja i profesionalne komunikacije (Al Daraai, Al Maqrashi i Al Shaikh, 2024; Chiu i sur., 2023; Cordona, Rodriguez i Ishmael, 2023; Mrnjaus, Vrcelj i Kušić, 2023; Okonkwo i Ade-Ibijola, 2021; World Economic Forum, 2024). Također, generativna umjetna inteligencija unaprjeđuje i znanstvene procese, pa tako ubrzava znanstvena otkrića analizom velikih skupova

podataka i povećava produktivnost znanstvenika smanjujući njihovo administrativno opterećenje. Također, znanstvenici koriste mogućnosti umjetne inteligencije za nadzirano učenje, simulaciju, predviđanje problema, kompresiju i interpretaciju podataka, utvrđivanje potencijalnih anomalija, istraživanje literature i upravljanje referencama, kreiranje hipoteza, lakše kvantificiranje nesigurnih rezultata istraživanja, automatiziranje rutinskih procesa i poslova, znanstvenu komunikaciju te pisanje rukopisa i pronalazak časopisa za objavu rukopisa (Arranz i sur., 2023; Bhosale, Phadke i Kapadia, 2023; Ghosh, 2023; Rainie i Anderson, 2024).

Dio istraživanja (Farhi i sur., 2023; Hogenhout i Takahashi, 2022; Zhang i Aslan, 2021) pokazuje da sve više studenata koristi generativnu umjetnu inteligenciju u svom obrazovanju. Najčešće ju koriste za jezično uređivanje tekstova i generiranje novih ideja, smatrajući da im umjetna inteligencija pomaže postići bolje obrazovne rezultate i uspjeh. Većina studenata podupire ideju integracije umjetne inteligencije u obrazovanje, no svjesni su i rizika, poput sve češćeg plagiranja, ugroze privatnosti podataka i, trenutno, nejasnih politika fakulteta i sveučilišta kada je u pitanju umjetna inteligencija (Almassaad, Alajlan i Alebaikan, 2024; Arowosegbe, Alqahtani i Oyelade, 2024; Chan i Hu, 2023; Mahmud i sur., 2024).

Kada su u pitanju izazovi i rizici, nastavnici i znanstvenici su često oprezniji od studenata. Svjesni su izazova i etičkih implikacija koje donosi nagla integracija umjetne inteligencije. Najviše ih brine akademsko nepoštenje, općenito pad akademskih vještina kod studenata, posebice pad kritičkog mišljenja, ali i vlastiti nedostatak vještina potrebnih za rad s alatima umjetne inteligencije. Većina je nastavnika i znanstvenika svjesna i prepoznaje mogućnosti umjetne inteligencije, posebno u kontekstu personaliziranog učenja, administrativne učinkovitosti te poboljšanja i ubrzanja istraživačkih procesa, no upozoravaju na nedostatak jasne komunikacije, jasnih smjernica i edukacije, što ponekad doprinosi stvaranju polarizirajućih narativa te lakom i čestom mijenjaju mišljenja i stavova (Bhosale, Phadke i Kapadia, 2023; Fadlelmula i Qadhi, 2024; Gerlich, 2023; McGrath i sur., 2023; Oyetola, Oladokun i Dogara, 2024; Zhang i Aslan, 2021).

Unatoč ozbiljnim etičkim izazovima, od pitanja plagijata, preko ugroze privatnosti, do smanjenja ljudske autonomije, generativna umjetna inteligencija doista mijenja i, kako će neki reći, revolucionira visoko obrazovanje i znanost kroz spomenute trendove personaliziranog učenja, novih pedagoških modela i brže znanstvene produkcije. Zbog toga je važno uspostaviti uravnotežen pristup integraciji umjetne inteligencije, što posebno uključuje razvoj jasnih i fleksibilnih obrazovnih politika u kojem sudjeluju svi dionici obrazovnog procesa te osigurati kontinuirano obrazovanje studenata i nastavnika za odgovorno korištenje umjetne inteligencije (Chan i Colloton, 2024; Cordona, Rodriguez i Ishmael, 2023; Zhang i Aslan, 2021).

Određeni autori (Chan i Colloton, 2024; Cordona, Rodriguez i Ishmael, 2023; Ghandour, 2024; Olędzka i sur., 2024; Yang i sur., 2021) naglašavaju humanocentrični pristup integraciji umjetne inteligencije u obrazovanje i znanost i činjenicu da bi umjetna inteligencija trebala imati ulogu asistenta studentima i nastavnicima, a ne zamijeniti njihov originalni intelektualni i kreativni rad. Upravo zbog toga promicanje pismenosti o umjetnoj inteligenciji (dalje: pismenost o UI-u) (engl. *AI literac*y) postaje nužnim prvim korakom. Ako se pravilno razumiju mogućnosti, ograničenja i etičke implikacije korištenja umjetne inteligencije te ako se ona koristi odgovorno i transparentno, onda uistinu može obogatiti iskustvo učenja i poučavanja, smanjiti nejednakosti i uspješno pripremiti studente za zanimanja i tržište rada budućnosti. No bez jasnih pravila i smjernica, posebice po pitanju etičkih implikacija, postoji rizik od produbljivanja postojećih obrazovnih, time i društvenih problema, a posebno narušavanja akademskog integriteta i čestitosti i pada povjerenja u visokoškolske institucije i akademsku zajednicu.

4.3. Etički aspekti integracije umjetne inteligencije u visoko obrazovanje i znanost

Posljednjih je nekoliko godina generativna umjetna inteligencija, s alatima poput ChatGPT-a, Perplexityja, Consesusa i drugih, postala nezaobilazna tema u visokom obrazovanju i znanosti. Uz spomenute prednosti, otvaraju i niz novih pitanja, od kojih su neka od ključnih: „Hoće li integracija umjetne inteligencije u obrazovanje i znanost umanjiti vrijednost ljudske kreativnosti i rada?“, „Kako zaštiti privatnost podataka?“ i „Kako očuvati akademski integritet i čestitost?“ (Mrnjaus, Vrcelj i Kušić, 2023; Sabzalieva i Valentini, 2024). Upravo potonje čini glavnu okosnicu ovoga rada.

Korištenje generativne umjetne inteligencije u obrazovanju i znanosti ne smije se promatrati samo kroz tehničku i tehnološku prizmu, već i kroz prizmu etičkih izazova koje nameće, od ugroze privatnosti podataka i nejednakosti u pristupu tehnologiji do pristranosti algoritama i diskriminacije. Stoga stručnjaci iz područja obrazovanja, znanosti, ali i samog područja umjetne inteligencije upozoravaju na nužnost razvoja etičkih okvira, smjernica i preporuka, kreiranja agilnih obrazovnih politika koje mogu relativno brzo odgovoriti na tehnološke promjene te razvoja nacionalne i međunarodne suradnje u akademskoj zajednici (Bhosale, Phadke i Kapadia, 2023).

Mnoga istraživanja (Arowosegbe, Alqahtani i Oyelade, 2024; Benouachane, 2024; Bhosale, Phadke i Kapadia, 2023; Chan i Colloton, 2024; Mahumud i sur., 2024; Isiaku i sur., 2024; Khatri i Karki, 2023; Kralj i sur., 2024; Mrnjaus, Vrcelj i Kušić, 2023; Msambwa, Wen i Daniel, 2025; Sabzalieva i Valentini, 2024) po-

kazuju da unatoč brojnim prednostima integracije umjetne inteligencije, postoje i značajni rizici od kojih su najčešći:

- plagiranje i varanje (predavanje pisanih radova i zadataka generiranih umjetnom inteligencijom kao vlastitih originalnih radova)
- nepouzdanost alata za prepoznavanje sadržaja generiranog umjetnom inteligencijom (alati su trenutačno nepouzdani i često daju lažne pozitivne ili lažne negativne rezultate)
- algoritamska pristranost (korištenje sustava umjetne inteligencije koji reproduciraju stereotipe i netočne informacije iz podataka na temelju kojih su trenirani)
- širenje dezinformacija (sustavi umjetne inteligencije mogu pružati pogrešne ili izmišljene informacije)
- ugrožavanje privatnosti podataka (alati i sustavi umjetne inteligencije obrađuju velike količine podataka, što otvara pitanje njihove sigurnosti)
- smanjenje ljudskih vještina (dugoročno korištenje alata i sustava umjetne inteligencije može stvorit pretjeranu ovisnost o njima i tako oslabiti vještine kritičkog mišljenja i kreativnosti kod studenata i nastavnika).

Tu su i dodatni izazovi koji uključuju razvoj tehnologije koji je brži od razvoja etičkih okvira, pravila i smjernica, česti otpor prema promjenama i nedostatak infrastrukture, resursa i edukacije, što otežava prilagodbu studenata i nastavnika novom obrazovnom okruženju (Al Daraai, Al Maqrashi i Al Shaikh, 2024; Fadlelmula i Qadhi, 2024; Miao i Holmes, 2023; Bhosale i sur., 2023; Sabzalieva i Valentini, 2023).

Nepouzdanost alata za prepoznavanje sadržaja generiranog umjetnom inteligencijom, možemo reći, dodatno komplicira situaciju. Jednostavnom usporedbom napisanih i generiranih tekstova više nije moguće s najvećim pouzdanjem utvrditi je li ih proizveo čovjek ili stroj. Iako su do pisanja ovoga rada razvijeni robusni alati za tu namjenu, koji su u stanju prepoznati tekstove koje je proizvela generativna umjetna inteligencija (npr. Originality.ai, GPTZero.me, ZeroGPT.com, itd.) – njihova pouzdanost još uvijek nije na zavidnoj razini (Khalil i Er, 2023; Liang, i sur., 2023). Štoviše, pokazalo se da autore kojima engleski nije materinski jezik – ti alati često „prokazuju“ kao plagijatore, a što naglašavaju Khalil i Er (2023) i Liang i sur. (2023). Rješenje tog problema danas se još uvijek ne nazire. Postavlja se pitanje: jesmo li s prodorom generativne umjetne inteligencije u akademsko okruženje – na dobitku ili na gubitku?

Kako odgovoriti na navedene etičke izazove? Prvi konkretniji odgovor akademskih institucija diljem svijeta na pojavu alata generativne umjetne inteligencije poput *chatbota* ChatGPT tvrtke OpenAI, bila je njihova zabrana. Ta-

kav je pristup, barem privremeno, usvojen na nekolicini visokoobrazovnih institucija, poput Sciences Po (Pariz) i Sveučilišta u Hong Kongu – da spomenemo samo ona koja su svoje odluke učinila javno dostupnima na internetu. Moramo istaknuti da je danas najkorišteniji *chatbot* na svijetu – OpenAI-ev ChatGPT – jedno kratko vrijeme (u travnju 2023.), zbog neusklađenosti s GDPR-legislativom, bio zabranjen u cijeloj Italiji.

Xiao, Chen i Bao (2023) u svom radu ističu da se alati umjetne inteligencije zabranjuju kao privremena mjera tijekom prve faze politike primjene umjetne inteligencije u obrazovanju. Isti rad izdvaja tri pristupa sveučilišta navedenom problemu: čekanje (engl. *waiting*) – većina institucija oklijeva s donošenjem jasnih smjernica, čekajući daljnji razvoj tehnologije ili konsenzus u akademskoj zajednici; zabranjivanje (engl. *banning*) – manji broj institucija uvodi stroga ograničenja korištenja; i prihvaćanje (engl. *embracing*) – manji broj institucija dopušta integraciju alata generativne umjetne inteligencije u nastavi, ali uz jasna pravila. Danas većina autora (Arowosegbe, Alqahtani i Oyelade, 2024; Bhosale, Phadke i Kapadia, 2023; Chan i Colloton, 2024; European Commission, 2024; Mahmud i sur., 2024; Miao i Holmes, 2023; OECD, 2019; Okonkwo i Ade-Ibijola, 2021; Rainie i Anderson, 2024; Sabzalieva i Valentini, 2023; World Economic Forum, 2024) naglašava potrebu za jasnim institucionalnim politikama koje će definirati dopuštene i nedopuštene svrhe korištenja umjetne inteligencije te uvođenje mjera kao što su transparentno korištenje alata umjetne inteligencije uz očuvanje ljudske autonomije, redizajn nastavnih sadržaja i zadataka tako da razvijaju kreativnost i kritičko razmišljanje kod studenata, aktivno uključivanje studenata i nastavnika u kreiranje institucionalnih politika kako bi se potaknulo, tj. povećalo razumijevanje i prihvaćanje pravila te uvođenje sustava za detekciju sadržaja generiranih umjetnom inteligencijom (unatoč činjenici da su gotovi svi takvi alati još uvijek nepouzdani). Dio mjera odnosi se i na promjenu paradigme u obrazovanju i znanosti. Velika je potreba za redizajnom postojećih kurikuluma i dizajnom novih koji će uključivati nove vještine i kompetencije, poput razumijevanja ograničenja algoritama i etičkog prosuđivanja njihovih odluka (Arowosegbe i sur., 2024; Chan i Colloton, 2024; Mrnjaus, Vrcelj i Kušić, 2023; Sabzalieva i Valentini, 2023; World Economic Forum, 2024). U kontekstu znanstvenih istraživanja, s obzirom na to da umjetna inteligencija sve više sudjeluje u prikupljanju i analizi istraživačkih podataka, nužni su robusni mehanizmi provjere kako bi se spriječili pristranost, ugroza privatnosti i širenje lažnih i netočnih informacija (Arranz i sur., 2023; Bhosale i sur., 2023; Ghosh, 2023). Dokument *Living guidelines on the responsible use of Generative AI in research* (European Commission, 2024) donosi tri skupine posebnih preporuka za znanstvenike, znanstvene ustanove i organizacije koje financiraju znanstvena istraživanja, a koje uključuju transparentno korištenje umjetne inteligencije, odgovornost i očuvanje akademskog integriteta i čestitosti, očuvanje privatnosti i osiguravanje prava intelektualnog vlasništva, pridržavanje

europskih zakonodavnih i etičkih okvira u području umjetne inteligencije, promicanje odgovornog korištenja umjetne inteligencije, i dr.

Generativna umjetna inteligencija predstavlja veliki tehnološki iskorak koji mijenja način na koji učimo, pišemo, istražujemo i stvaramo, no ujedno otvara složena etička pitanja koja ponajprije uključuju njezino odgovorno korištenje i očuvanje akademskog integriteta i čestitosti. Etika umjetne inteligencije u visokom obrazovanju i znanosti traži višedimenzionalni pristup koji povezuje napredne mogućnosti tehnologije s ljudskim vrijednostima i demokratskim načelima. Odgovorna integracija tehnologije umjetne inteligencije ne znači samo smanjivanje, ograničavanje i eliminiranje rizika već aktivno oblikovanje budućnosti obrazovanja i istraživanja prema načelima pravednosti, inkluzivnosti, transparentnosti i ljudske autonomije (European Commission, 2024). Sve češći interes organizacija kao što su UNESCO i određena europska regulatorna tijela, ali i pojedini fakulteti i sveučilišta za pitanje odnosa (generativne) umjetne inteligencije i akademskog integriteta i čestitosti indikator je da je upravo to postalo jednim od prioriteta etike umjetne inteligencije. Kreiranje jasnih smjernica više nije izbor nego nužnost kako bi se očuvali kvaliteta znanja, kultura inovacije i opće akademske vrijednosti.

5. Akademski integritet i etičke implikacije generativne umjetne inteligencije u visokom obrazovanju: važnost pismenosti o UI-u

U knjizi *Neksus*: *kratka povijest informacijskih mreža od kamenog doba do umjetne inteligencije*, objavljenoj 2024. godine, Yuval Noah Harari problematizira povjerenje u autorstvo pisanih sadržaja (Harari, 2024). Iako tvrdi da je knjigu napisao osobno, uz pomoć nekoliko suradnika, priznaje da čitatelji u to više ne mogu biti sigurni, za razliku od vremena prije nedavnog razvoja umjetne inteligencije. Taj primjer ilustrira veliku promjenu koja se pojavom generativne umjetne inteligencije dogodila na području, dosad rezerviranom isključivo za ljudsku kreaciju – akademskom pisanju. Ne samo da nas je generativna umjetna inteligencija iznenadila svojom visokom kvalitetom proizvedenih tekstova na bilo koju temu nego je otvorila vrata jednoj do sada nezamislivoj i neugodnoj mogućnosti – da znanstvenici i studenti potpisuju svoje radove iza kojih, zapravo, stoji umjetna inteligencija.

Korištenje ChatGPT-a i sličnih alata generativne umjetne inteligencije u visokom obrazovanju i znanosti uzburkalo je raspravu o akademskom integritetu i čestitosti postavivši posve nove izazove pred akademsku zajednicu. U početku su mnogi strahovali od „epidemije“ varanja među studentima, no ubrzo se ispostavilo da generativna umjetna inteligencija donosi i brojne prednosti poput spomenute personalizacije učenja, ubrzavanja procesa istraživanja, kreiranja inovativnih pedagoških praksi i smanjenja administrativnog opterećenja nastavnika. Neki autori

(Ali i sur., 2024) čak ističu pozitivan utjecaj generativne umjetne inteligencije na akademski integritet i čestitost vjerujući da otvoreno, transparentno, odgovorno i etičkim okvirima i smjernicama usmjereno korištenje tehnologije umjetne inteligencije može osnažiti i dodatno motivirati studente da se pridržavaju načela akademskog integriteta i čestitosti. Novi alati koji mogu generirati tekst nalik onomu koji je napisao čovjek, koji mogu rješavati složene zadatke, predlagati nove ideje i koncepte i nuditi personalizirane odgovore u nekoliko sekundi otvorili su posve nove mogućnosti za nastavnike i studente (Fadlelmula i Qadhi, 2024; Neves i sur., 2024; Okonkwo i Ade-Ibijola, 2021; Perkins, 2023; Sabzalieva i Valentini, 2024). Ipak briga i strah od nove tehnološke transformacije sustava obrazovanja i znanosti potaknuli su brojna pitanja. „Kako sačuvati akademski integritet i čestitost u doba u kojem strojevi mogu pisati eseje poput čovjeka?“, „Kako redefinirati pojam „originalnosti“ u obrazovanju i znanosti?“, „Hoće li pretjerano korištenje tehnologije umjetne inteligencije negativno utjecati na razvoj kritičkog mišljenja i sposobnost samostalnog rješavanja problema?“.

U tom smislu pojavio se i pojam *UI-giranje* (engl. *AI-giarism*). Premda prvi put spomenut u listopadu 2022. godine, samo mjesec dana prije objave ChatGPT-a, i to u jednoj objavi na tadašnjem Twitteru, ubrzo je prepoznat kao ključni pojam u širim raspravama o akademskom integritetu i čestitosti. U znanstvenom kontekstu, pojam je detaljnije razrađen u radu *is ai changing the rules of academic misconduct? an in-depth look at students' perceptions of 'AI-giarism'* (Chan, 2023). Taj rad značajan je ne samo po tome što je u znanstvenu raspravu uveo nov koncept već i što predstavlja i prvo sustavno istraživanje njegove percepcije kod studenata. Rezultati su pokazali da se velika većina studenata protivi radovima koji su u potpunosti generirani umjetnom inteligencijom, ali i da postoje tzv. „sive zone“ za djelomičnu uporabu. Te „sive zone“ odnose se na korištenje alata generativne umjetne inteligencije za pomoćne zadatke poput lekture, parafraziranja dijelova drugih članaka, prijevoda, ali i generiranja ideja ili preoblikovanja rečenica (uređivanje), gdje studenti nisu sigurni smatra li se takva uporaba nepoštenom ili prihvatljivom. Sve to upućuje na potrebu za jasnijim smjernicama i edukacijom o granicama akademske čestitosti kako bi se izbjegao *UI-giranje* u akademskom pisanju.

Ponekad su stavovi i odnos prema generativnoj umjetnoj inteligenciji dosta složeni i dvojni (Gerlich, 2023). Iako studenti često naglašavaju praktične koristi poput bržeg učenja, veće učinkovitosti i kreativnosti, ujedno su zabrinuti zbog potencijalne netočnosti informacija i gubitka autonomije, odnosno smanjenja osobnog angažmana u procesu učenja. Nastavnici, pak, više naglašavaju izazove i rizike, posebno rizik povećanog plagiranja i smanjenja kritičkog mišljenja kod studenata. No, i jedni i drugi slažu se da u odnosu na složenost i izazovnost situacije u kojoj se nalaze visoko obrazovanje i znanost nedostaju jasne institucionalne politike i smjernice. Podjele stavova o korištenju generativne umjetne inteligenci-

je odnose se i na razlike u percepciji studenata po pitanju samog narušavanja akademskog integriteta i čestitosti, pa tako pisanje cijeloga rada uz pomoć umjetne inteligencije većina studenata smatra težom povredom akademskog integriteta, dok manju pomoć umjetne inteligenciju u radu na nekom zadatku smatraju lakšom povredom (Lund i sur., 2025). Također studenti se suočavaju i s problemom podijeljenih očekivanja jer poslodavci sve više traže razvijene vještine korištenja umjetne inteligencije, dok veliki broj fakulteta još uvijek zabranjuje ili ograničava korištenje umjetne inteligencije i time usporava razvoj potrebnih vještina kod studenata (Felicity, 2024).

U nedavno provedenom istraživanju časopisa *Nature* vezanom za praksu korištenja alata generativne umjetne inteligencije u akademskom okruženju, preko 90 % ispitanika smatra prihvatljivim korištenje navedenih alata za lekturu, jezičnu ispravku i prijevod članaka, dok je oko dvije trećine ispitanika suglasno s uporabom tih alata za izradu prvog nacrta ili određenih dijelova rukopisa. Jednu zanimljivost predstavlja otkriće da se alati generativne umjetne inteligencije u praksi koriste manje nego što je dopušteno – samo 28 % ispitanika doista koristi alate generativne umjetne inteligencije za lekturu, a svega 4 % za pisanje prvog nacrta ili recenziju. To ukazuje na stanovit oprez autora pri uporabi tih alata (Kwon, 2025).

Među najčešćim rizicima koji mogu narušiti akademski integritet i čestitost navode se plagiranje i tzv. autori duhovi (engl. *ghostwriting*), pristranost algoritama i širenje dezinformacija, smanjenje vještina kod studenata (posebice kritičkog mišljenja i kreativnosti) i narušavanje privatnosti (Bittle i El-Gayar, 2025; Fowler, 2023; Miao i Holmes, 2023; Neves i sur., 2024; Rasul i sur., 2024; Sabzalieva i Valentini, 2024; Song, 2024). Čest su problem i podaci na kojima se treniraju alati generativne umjetne inteligencije, a koji mogu dovesti do spomenutih problema pristranosti i dezinformacija. U nedavno objavljenom radu Wilson i Burleigh (2025) dokumentiraju izazove i rizike za akademski integritet i čestitost u znanosti – svi vezani za mogućnost zlouporabe korištenja alata generativne umjetne inteligencije. Istaknuto je i prikazano nekoliko primjera poput fabrikacije i lažiranja istraživačkih rezultata, nejasnog autorstva i pripisivanja doprinosa. Posebno je istaknuta potreba za transparentnošću u korištenju alata umjetne inteligencije s naglaskom na jasno navođenje upotrebe umjetne inteligencije u svim fazama istraživanja uz poticanje stalnog dijaloga među znanstvenicima, stručnjacima i donositeljima odluka. Također tehnologija generativne umjetne inteligencije samo je produbila postojeće probleme kao što su kupovina članaka, kupovina autorstva na člancima, kupovina citiranosti, i dr., dok je tzv. *Deepfake*-tehnologija omogućila manipulaciju rezultatima znanstvenih eksperimenata (Miao i Holmes, 2023).

Kao odgovor na naveden rizike, stručnjaci sve češće naglašavaju razvoj proaktivnih i sveobuhvatnih politika i strategija i kreiranje institucionalnih smjernica

koje bi trebale jasno definirati dopuštene i zabranjene načine korištenja umjetne inteligencije u obrazovanju i znanosti. Ključne preporuke uključuju:

- razvoj etičkih okvira, institucionalnih strategija i pravila za korištenje tehnologije umjetne inteligencije
- planiranje i sustavnu implementaciju umjetne inteligencije u proces učenja poučavanja i znanstveno-istraživačkog rada
- odgovorno i transparentno korištenje alata i sustava umjetne inteligencije
- kontinuirano praćenje i evaluaciju utjecaja umjetne inteligencije na proces učenja poučavanja i znanstveno-istraživačkog rada
- kontinuiranu edukaciju studenata i nastavnika i njihovo opismenjavanje o UI-u
- njegovanje i promicanje digitalne jednakosti među studentima
- inovaciju pedagoških praksi koje se temelje više na projektnom učenju, zadacima usmjerenima na primjenu kritičkog mišljenja i kreativnosti te na usmenom ispitu
- prilagodbu metoda vrjednovanja s naglaskom na vještine i kompetencije koje umjetna inteligencija ne može ili teško može reproducirati
- korištenje alata za prepoznavanje sadržaja generiranog umjetnom inteligencijom (uz svijest o njihovim ograničenjima)
- angažiranje etičkih odbora i skupina koje će činiti i osobe pismene u području UI-a
- uključivanje svih dionika visokog obrazovanja i znanosti, od mikro do makro razine (studenata, nastavnika, znanstvenika, uprave fakulteta i sveučilišta, političara i kreatora politika) u proces oblikovanja politika, smjernica i preporuka (Bhosale, Phadke i Kapadia, 2023; Castelló-Sirvent, Roger-Monzó i Gouveia-Rodrigues, 2024; Davis, 2024; Evangelista, 2025; Ghandour, 2024; Lund i sur., 2025; Miao i Holmes, 2023: Neves i sur., 2024; Perkins, 2023; Rasul i sur., 2024; Sabzalieva i Valentini, 2024).

Davis (2024) posebno naglašava edukaciju studenata te upozorava na određene skupine studenata (strani studenti, neurodivergentni studenti, studenti pripadnici manjina, i dr.) koje imaju poteškoća s razumijevanjem i slijeđenjem pravila akademskog integriteta i čestitosti kao na one na koje je potrebno obratiti posebnu pozornost u provođenju edukacija. Postoje i određeni praktični pristupi, poput ljestvice procjene korištenja umjetne inteligencije (engl. *AI Assessment Scale*) koja nudi prilično jednostavan način za kategorizaciju i procjenu razine korištenja umjetne inteligencije u visokom obrazovanju i znanosti (Perkins i sur., 2024).

Da bi mogli kritički koristiti alate generativne umjetne inteligencije, studenti i nastavnici moraju razviti vještine pismenosti u području umjetne inteligencije, odnosno razumjeti kako prednosti tako i ograničenja tehnologije umjetne inteligencije i etički procijeniti alate i sustave koje koriste. Pismenost o UI-u glavni je preduvjet očuvanja akademskog integriteta i čestitosti u novim uvjetima visokog obrazovanja i znanosti. Ona ne pretpostavlja samo tehničke aspekte i vještine korištenja umjetne inteligencije već i razumijevanje kako umjetna inteligencija funkcionira, kritičku procjenu informacija i sadržaja koje generira te svjesno promišljanje o etičkim implikacijama njezina korištenja (Chan i Colloton, 2024; Hornberger, Bewersdorff i Nerdel, 2023; Southworth i sur., 2023; Yi, 2021). Ng i suradnici (2021) ističu četiri temeljna aspekta pismenosti o UI-u iz kojih proizlazi prethodno spomenuta činjenica da se ne radi samo o tehničkoj vještini, a to su: 1) poznavanje i razumijevanje tehnologije umjetne inteligencije, 2) učinkovita primjena alata umjetne inteligencije u različitim kontekstima, 3) sposobnost stvaranja sadržaja uz pomoć umjetne inteligencije i kritička evaluacija istih te 4) etičko promišljanje i odgovorno korištenje umjetne inteligencije. Long i Magerko (2020) ističu, pak, tri osnovne dimenzije pismenosti o UI-u: 1) algoritamsku pismenost (razumijevanje kako algoritmi donose odluke), 2) podatkovnu pismenost (znanje o prikupljanju, obradi i korištenju podataka) i 3) etiku umjetne inteligencije. Dodatno naglašavaju i važnost metakognicije, tj. sposobnosti kritičkog promišljanja o vlastitim postupcima i predviđanja tehnoloških promjena koja pomaže prepoznati pravi utjecaj tehnologije umjetne inteligencije.

Mnogi su pogled uprli upravo prema (visokoškolskim) knjižnicama od kojih očekuju glavnu inicijativu u promicanju i razvoju pismenosti o UI-u, posebice s obzirom na njihovu ulogu i iskustvo s informacijskim i medijskim opismenjavanjem. Primjerice, knjižnice bi mogle pomoći nastavnicima u ažuriranju postojećih kurikuluma i edukaciji studenata i nastavnika za bolje razumijevanje rada alata umjetne inteligencije i algoritama, kao i same etike umjetne inteligencije (Walter, 2024).

Da se pismenost o UI-u često veže za informacijsku i medijsku pismenost svjedoči i prijedlog Ndungua (2024) o integraciji pismenosti u području umjetne inteligencije u okvire poput ACRL (engl. Association of College and Research Libraries) okvira za informacijsku pismenost u visokom obrazovanju (engl. *ACRL Framework for Information Literacy for Higher Education*). Slično povezivanje pismenosti u području umjetne inteligencije i informacijske pismenosti predlažu i Southworth i suradnici (2023). Većina autora (Hornberger, Bewersdorff i Nerdel, 2023; Long i Magerko, 2020; Miao i sur. 2021; Ng i sur., 2021; Southworth i sur., 2023; Yi, 2021) se slaže da etika mora biti središnji dio opismenjavanja i obrazovanja o umjetnoj inteligenciji. To uključuje kompetencije poput razumijevanja kako umjetna inteligencija donosi odluke (transparentnost), razumijevanja tko snosi odgovornost kada umjetna inteligencija pogriješi (odgovornost), izbje-

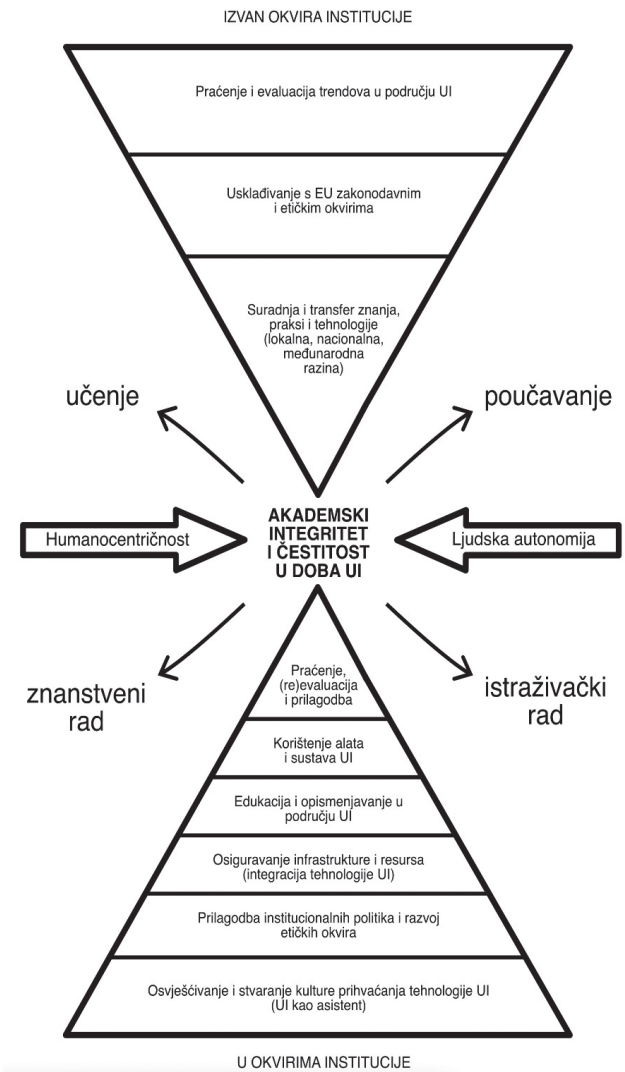
gavanja diskriminacije i digitalne nejednakosti (pravednost i inkluzija), prepoznavanja i ispravljanja predrasuda algoritama i podataka (ublažavanje pristranosti) i sigurnog upravljanja osobnim podacima (zaštita privatnosti). Bez navedenih etičkih kompetencija pismenost o UI-u nije potpuna i može samo produbiti postojeće izazove i probleme. Iako postoji konsenzus o važnosti pismenosti o UI-u, većina autora (Long i Magerko, 2020; Luckin i sur., 2022; Ndungu, 2024; Ng i sur., 2021; Yi, 2021) naglašava nekoliko izazova za njezinu širu implementaciju – nedostatak jedinstvenog okvira i jasnih definicija pojmova, nedovoljna stručnost studenata i nastavnika u području umjetne inteligencije, ograničeni resursi za edukaciju studenata i nastavnika i za kreiranje edukativnih materijala te rizik od produbljivanja digitalne nejednakosti među studentima. S obzirom na tehnološke trendove, kritičko razumijevanje i odgovorno korištenje tehnologije umjetne inteligencije postaju ključnim vještinama 21. stoljeća, stoga se uvođenje osnovnih znanja o umjetnoj inteligenciji preporučuje već u ranoj obrazovnoj dobi kako bi se mlade generacije što prije počele pripremati za budućnost rada i života u visoko tehnologiziranom društvu (Kong, Cheung i Zhang, 2021; Song, 2024; World Economic Forum, 2024).

Uz pismenost o UI-u, neki autori (Luckin i sur., 2022) predlažu i koncept „spremnosti za umjetnu inteligenciju“ (engl. *AI readiness*). Radi se o strukturiranom okviru za razumijevanje i primjenu umjetne inteligencije u obrazovanju, ali i drugim sektorima. Okvir uzima u obzir specifične potrebe različitih zanimanja i ističe netehnički pristup kako bi se osnažilo studente i nastavnike za odgovorno korištenje alata i sustava umjetne inteligencije. Kako se pristup i korištenje umjetne inteligencije razlikuje u pojedinim znanstvenim područjima (medicini, arhitekturi ili društvenim i humanističkim znanostima), tako je potrebno posebno prilagoditi kurikulume i pedagoške metode u svakom području. Na svojevrsni način „spremnost za umjetnu inteligenciju“ podrazumijeva širu kontekstualizaciju pismenosti o UI-u i njezinu prilagodbu za svaku pojedinu profesiju.

Na temelju svega navedenog, može se izvesti zaključak sličan onome do kojeg su došli i Bittle i El-Gayar (2025) u svom radu, ističući da alati generativne umjetne inteligencije, s jedne strane mogu poboljšati angažman i učinkovitost studenata i znanstvenika u visokom obrazovanju, ali da, s druge strane, nose rizik nepoštivanja akademske čestitosti i narušavanja akademskog integriteta. Iznalazak ravnoteže između navedenih dvaju ključnih obilježja aspekata generativne umjetne inteligencije koji se odnose na visoko obrazovanje predstavlja ključ prilagodbe budućih kurikuluma u dobu koje se pomalja na obzoru, i koje će obilježiti era tehnologije koja nadilazi sve prethodne ljudske tvorevine – umjetne inteligencije.

6. Umjesto zaključka – Prijedlog konceptualnog okvira i smjernica za očuvanje akademskog integriteta i čestitosti u doba umjetne inteligencije

Predloženi konceptualni model (slika 1) prikazuje okvir za očuvanje akademskog integriteta i čestitosti u doba umjetne inteligencije povezujući institucionalne odgovornosti sa širim izvaninstitucionalnim kontekstom.



Slika 1. Konceptualni okvir za očuvanje akademskog integriteta i čestitosti u doba umjetne inteligencije.

U središtu modela nalaze se **akademski integritet i čestitost**, koji predstavljaju etički temelj svih procesa **učenja**, **poučavanja**, **znanstvenog i istraživačkog rada**. Dva ključna načela, **humanocentrični** pristup i načelo očuvanja **ljudske autonomije**, usmjeravaju integraciju umjetne inteligencije u kontekstu akademskog integriteta i čestitosti tako što osnažuju sve dionike u sustavu visokog obrazovanja i znanosti, umjesto da umanjuju ili zamjenjuju njihove mogućnosti.

Gornji trokut prikazuje vanjske (izvaninstitucionalne) čimbenike koji utječu na oblikovanje i očuvanje institucionalne kulture akademskog integriteta i čestitosti u doba umjetne inteligencije. Radi se o čimbenicima koji nisu dio ustroja i djelovanja institucije, no koji (ne)izravno utječu na nju i njezin rad. Proces započinje **praćenjem i evaluacijom trendova u području umjetne inteligencije** koji omogućuju informirano i strateško planiranje i integraciju tehnologije umjetne inteligencije u visokoškolske ustanove. Potom slijedi **usklađivanje s europskim zakonodavnim i etičkim okvirima** o umjetnoj inteligenciji općenito i u sustavu visokog obrazovanja i znanosti konkretno, u sklopu kojeg se osigurava usklađenost institucionalnih etičkih okvira s aktualnim europskim regulativama. Posljednji korak uključuje **suradnju i transfer znanja, praksi i tehnologija**[1] na lokalnoj, nacionalnoj i međunarodnoj razini. Razmjena najboljih praksi i iskustava potiče zajednički etički pristup integraciji umjetne inteligencije u procese učenja, poučavanja i znanstveno-istraživačkog rada i osigurava interoperabilnost standarda akademskog integriteta i čestitosti.

Donji trokut prikazuje unutarnje (institucionalne) čimbenike i definira institucionalne mjere nužne za izgradnju i očuvanje kulture akademskog integriteta i čestitosti. Prvi korak obuhvaća **osvješćivanje** svih unutarnjih dionika (studenti, nastavnici, uprava) **i stvaranje kulture prihvaćanja tehnologije umjetne inteligencije** u kojoj ona ima ulogu asistenta koji osnažuje navedene dionike[2]. Potom slijedi **redizajn institucionalnih politika** koje moraju biti jasne, inkluzivne i fleksibilne **i razvoj etičkih okvira** kojima je cilj unaprijediti institucionalnu politiku kvalitete, rad etičkih povjerenstava i povezanih povjerenstava i radnih skupina te etički kodeks i ostale povezane dokumente kako bi institucija što lakše i uspješnije

[1] Primjerice, transfer lokalne tehnologije razvijene na institucionalnoj razini (specifični alati i sustavi umjetne inteligencije, specifični agenti umjetne inteligencije i sl.), transfer samih etičkih okvira (smjernica, preporuka, načela, i dr.), transfer pedagoških praksi i nastavnih sadržaja, i sl. koji se mogu primijeniti i u drugim institucionalnim kontekstima, posebice onima kojima nedostaju infrastruktura i resursi.

[2] Za kreiranje učinkovitih institucionalnih etičkih okvira i smjernica i očuvanje akademskog integriteta i čestitosti nužna je podrška cjelokupnog sustava visokog obrazovanja i znanosti koji je dovoljno osviješten i informiran o mogućnostima i prilikama, ali i izazovima i rizicima tehnologije umjetne inteligencije. To podrazumijeva osvješćivanje i prihvaćanje tehnologije umjetne inteligencije i među vanjskim dionicima kao što su kreatori (obrazovnih) politika, donositelji odluka, nadležno ministarstvo, i dr.

pratila trendove i promjene koje nameće razvoj tehnologije umjetne inteligencije. Također je važno razviti vlastiti etički okvir[3] (mjere, pravila, smjernice) po pitanju integracije i korištenja alata i sustava umjetne inteligencije, a pri njegovu oblikovanju važno je uključiti sve unutarnje dionike kako bi se osigurali legitimnost i prihvaćanje pravila. Sljedeći je korak **osiguravanje infrastrukture i resursa (integracija tehnologije umjetne inteligencije)**, a podrazumijeva osiguranje internet veze, opreme, alata i sustava umjetne inteligencije, licenci, i dr.) kod kojeg je važno voditi se načelom inkluzivnosti, odnosno osigurati pristup infrastrukturi i resursima svim dionicima. Nakon toga slijede **edukacija i opismenjavanje o UI-u**. U odnosu na kontinuirani i brzi razvoj tehnologije umjetne inteligencije, potrebno je periodično prilagođavati kurikulum, redizajnirati nastavne sadržaje i inovirati pedagoške prakse kako bi studenti stekli potrebna specifična znanja i vještine za rad s alatima i sustavima umjetne inteligencije u kontekstu vlastite struke, s naglaskom na kritičkom mišljenju i kreativnosti koji onemogućuju ovisnost o tehnologiji. Također, kroz oblikovanje posebnih kolegija ili kroz integraciju umjetne inteligencije u postojeće kolegije i kroz programe stručnog usavršavanja, važno je uključiti opću pismenost o UI-u za sve dionike. U tom je koraku važna uloga visokoškolske knjižnice i knjižničara koji se, zahvaljujući znanju i iskustvu u području informacijske, medijske, digitalne i drugih pismenosti, mogu aktivno uključiti u proces edukacije i opismenjavanja. Peti korak obuhvaća aktivno **korištenje alata i sustava umjetne inteligencije**. Pritom je važno poštivati načela transparentnosti i odgovornosti kako bi se održalo povjerenje i pravednost među dionicima. U konačnici, potrebno je kontinuirano **pratiti, (re)evaluirati i, po potrebi, prilagođavati korištenje umjetne inteligencije** kako bi se uvijek osiguralo kritičko i etički odgovorno korištenje alata i sustava umjetne inteligencije. U tom smislu, važan je rad etičkih odbora[4] čiji su članovi pismeni o UI-u.

Važno je napomenuti na predloženi konceptualni okvir i mjere zagovaraju pristup prihvaćanja tehnologije umjetne inteligencije, umjesto čekanja i zabrane, pri čemu umjetna inteligencija predstavlja asistenta koji osnažuje sve dionike u susta-

[3] Pravila mogu i trebaju biti što konkretnija kako bi se izbjegle dvosmislenosti u njihovu tumačenju. Primjerice, to može uključivati točno određivanje dozvoljenog i nedozvoljenog korištenja umjetne inteligencije, utvrđivanje konkretnih mjera za kršenje akademskog integriteta i čestitosti korištenjem umjetne inteligencije, formalne izjave o korištenju umjetne inteligencije koje bi se prilagale pisanim radovima, prilagodbu postojećih sustava citiranja koji bi uključivali i umjetnu inteligenciju kao jedan oblik „izvora“, i sl.

[4] Jedno od zanimljivih pitanja koje se nameće jest hoće li postupnom automatizacijom administrativnih i rutinskih zadataka, a u kontekstu razvoja alata umjetne inteligencija za detekciju plagijata i drugih oblika prijevara, doći i do potpune automatizacije procesa procjene akademskog integriteta (koji se velikim dijelom oslanja upravo na objektivna pravila i mjere podložne automatizaciji). Time bismo došli u situaciju da umjetna inteligencija na kraju prosuđuje i presuđuje samoj sebi. To pitanje možda još više naglašava potrebu za očuvanjem ljudskog nadzora i autonomije unatoč tehnološkom napretku.

vu visokog obrazovanja i znanosti u njihovom profesionalnom radu. To je posebice važno kada su u pitanju studenti čija (buduća) struka i (buduće) tržište rada očekuju razvijene znanja i vještine u području (generativne) umjetne inteligencije.

Integracija generativne umjetne inteligencije u visoko obrazovanje i znanost predstavlja jedno od najznačajnijih etičkih i pedagoških izazova suvremenog doba. Brzi razvoj tehnologije umjetne inteligencije donosi iznimne mogućnosti za procese učenja, poučavanja i znanstveno-istraživačkog rada osnažujući intelektualni i kreativni potencijal studenata i nastavnika, ali i otvarajući složena pitanja očuvanja akademskog integriteta i čestitosti. Spomenuti etički izazovi zahtijevaju sustavne i proaktivne politike i mjere kako bi se spriječile zlouporabe i osigurala kritička, transparentna i odgovorna primjena tehnologije.

Uloga umjetne inteligencija kao asistenta u radu studenata, nastavnika i znanstvenika nije „samo" izazov za pitanje akademskog integriteta i čestitosti *per se*, već naglašava i potrebu za redefiniranjem pojma originalnosti i kreativnosti. Još jedna važna činjenica koju je potrebno uzeti u obzir pri oblikovanju obrazovnih i znanstvenih politika i etičkih okvira jest ta da suvremeni visokoškolski sustav oblikovan trendovima tehnologije umjetne inteligencije nikada nije tražio veći naglasak na interdisciplinarnosti i konvergenciji znanja, vještina i kompetencija.

Odgovorna integracija generativne umjetne inteligencije u sustav visokog obrazovanja i znanosti može značajno unaprijediti kvalitetu obrazovanja, znanstvenog rada i istraživanja, pod uvjetom da se pitanje etičkih implikacija postavi u središte rasprave. Očuvanje akademskog integriteta i čestitosti u tom procesu ne znači samo sprječavanje nepoštenja, već i poticanje kulture transparentnosti, kritičkog promišljanja i odgovornosti koja će akademsku zajednicu pripremiti za izazove i prilike budućnosti.

LITERATURA

Adams, F. (2003). The Informational turn in philosophy. *Minds and Machines* 13, 4: 471–501. https://doi.org/10.1023/A:1026244616112

Al Daraai, S. B.; M. Al Maqrashi i Z. Al Shaikh (2024). Integrating AI in higher education: applications, strategies, ethical considerations. U: N. H. Al Harrasi i M. S. El Din (ur.) *Utilizing AI for Assessment, Grading, and Feedback in Higher Education* (str. 189–211). Hershey PA: IGI Global.

Ali i sur. (2024). Ali, W.; R. Alami; M. A. K. Alsmairat i AlMasaeid, T. Consensus or controversy: examining AI's impact on academic integrity, student learning, and inclusivity within higher education environments. U: *2024 2nd International Conference on Cyber Resilience (ICCR)* (str. 1–5). Dubai, United Arab Emirates: Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCR61006.2024.10532968

Almassaad, A.; H. Alajlan i R. Alebaikan (2024). Student perceptions of Generative Artificial Intelligence: investigating utilization, benefits, and challenges in higher education. *Systems* 12: 385. https://doi.org/10.3390/systems12100385

Arowosegbe, A.; J. S. Alqahtani i T. Oyelade (2024) Perception of Generative AI use in UK higher education. *Frontiers in Education* 9:1463208. doi: 10.3389/feduc.2024.1463208

Arranz i sur. (2023). Arranz, D.; Bianchini, S.; Di Girolamo, B. i Ravet, J. *Trends in the use of AI in science: a bibliometric analysis*. Brussels: European Commission. [citirano: 2025–07–02]. Dostupno na: https://op.europa.eu/en/publication-detail/-/publication/2458267c-08df-11ee-b12e-01aa75ed71a1

Benouachane, H. (2024). AI in higher education: balancing pedagogical benefits and ethical challenges. *Educational AI Review* 2, 5: 302–322. doi: 10.6084/m9.figshare.26349289

Biever, C. (2023). ChatGPT broke the Turing test — the race is on for new ways to assess AI. Nature. [citirano: 2025–07–13]. Dostupno na: https://www.nature.com/articles/d41586-023-02361-7

Bittle, K. i O. El-Gayar (2025). Generative AI and academic integrity in higher education: a systematic review and research agenda. *Information 16,* 4: 296. https://doi.org/10.3390/info16040296

Bhosale, U.; G. Phadke i A. Kapadia (2023). *Generative AI in research: a practical guide for universities on balancing risks and benefits*. *Enago*. [citirano: 2025–07–02]. Dostupno na: https://kc.umn.ac.id/id/eprint/28037/

Borenstein, J. i A. Howard (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics* 1: 61–65. https://doi.org/10.1007/s43681-020-00002-7

Bosančić, B. (2023). *Informacija u teoriji*. Zagreb: Naklada Ljevak.

Bostrom, N. (2014) *Superintelligence: paths, dangers, strategies.* Oxford: Oxford University Press.

Bubeck i sur. (2023). Bubeck, S.; V. Chandrasekaran; R. Eldan; J. Gehrke; E. Horvitz; E. Kamar, P. Lee i suradnici. Sparks of Artificial General Intelligence: early experiments with GPT-4. *ArXiv:2303.12712*. https://arxiv.org/abs/2303.12712

Bush, V. (1945). As we may think. *The Atlantic Monthly* 176, 1: 101–108. [citirano: 2025–07–02]. Dostupno na: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

Castelló-Sirvent, F.; V. I. Roger-Monzó i R. Gouveia-Rodrigues (2024). Quo vadis, university? a roadmap for AI and ethics in higher education. *The Electronic Journal of e-Learning: Special Issue: Artificial Intelligence (AI) in Education. Part 2.* 22, 6: 34–51. https://doi.org/10.34190/ejel.22.6.3267

Castillo, E. (2023). *These schools and colleges have banned ChatGPT and similar AI tools*, BestColleges. [citirano: 2025–07–25]. Dostupno na: https://www.bestcolleges.com/news/schools-colleges-banned-chat-gpt-similar-ai-tools/

Chan, C. K. Y. (2023). Is AI changing the rules of academic misconduct? An in-depth look at students' perceptions of'AI-giarism'. *arXiv preprint arXiv:2306.03358*. https://doi.org/10.48550/arXiv.2306.03358

Chan, C. K. Y. i W. Hu (2023). Students' voices on Generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20: 43. https://doi.org/10.1186/s41239-023-00411-8

Chan, C. K. Y. i Colloton, T. (2024). *Generative AI in higher education: the ChatGPT effect*. London, New York: Routledge.

Chiu i sur. (2023). Chiu, T. K. F.; Q. Xia; X. Zhou; C. S. Chai i M. Cheng. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence* 4: 100118. https://doi.org/10.1016/j.caeai.2022.100118

Coeckelbergh, M. (2020). *AI ethics*. Cambridge: The MIT Press.

Cordona, M. A.; Rodriguez, J. R. i Ishmael, K. (2023). *Artificial Intelligence and the future of teaching and learning: insights and recommendations*. Washington, DC: Office of Educational Technology. [citirano: 2025–07–02]. Dostupno na: https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Davis, M. (2024). Supporting inclusion in academic integrity in the age of GenAI. U: Beckingham S.; J. Lawrence; S. Powell i P. Hartely (ur.) *Using Generative AI Effectively in higher education: sustainable and ethical practices for learning, teaching and assessment* (str. 21–30). London, New York: Rourledge.

European Commission. (2024). *Living guidelines on the responsible use of Generative AI in research*. Brussels: European Union. [citirano: 2025–07–02]. Dostupno na: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

European Commission. (2021). *Study to support an impact assessment of regulatory requirements for artificial intelligence in Europe: final report*. Brussels: European Union. [citirano: 2025–07–02]. Dostupno na: https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1/language-en

Europska komisija. (2020). *Bijela knjiga o umjetnoj inteligenciji: europski pristup izvrsnosti i izgradnji povjerenja*. Brussels: Europska komisija. [citirano: 2025–07–02]. Dostupno na: https://op.europa.eu/hr/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1

European Union. (2024). *Artificial Intelligence Act*. Brussels: European Union. [citirano: 2025–07–02]. Dostupno na: https://artificialintelligenceact.eu/the-act/

Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology* 17*,* 1: ep559. https://doi.org/10.30935/cedtech/15775

Fadlelmula, F. K. i Qadhi, F. K. (2024). A systematic review of research on artificial intelligence in higher education: practice, gaps and future directions in the GCC.

Journal of University Teaching and Learning Practice 21, 6: 146–173. https://doi.org/10.53761/pswgbw82

Felicity, T. (2024). College students struggle to balance AI use with academic integrity policies set by universities. *University Herald*. [citirano: 2025–07–02]. Dostupno na: https://www.universityherald.com/articles/79446/20241216/college-students-struggle-balance-ai-use-academic-integrity-policies-set-universities.htm#:~:text=For%20students%2C%20learning%20AI%20feels,how%20to%20honor%20academic%20integrity

Fowler, D. S. (2023). AI in higher education: academic integrity, harmony of insights, and recommendations. *Journal of Ethics in Higher Education* 3: 127–143. https://doi.org/10.26034/fr.jehe.2023.4657

Gerlich, M. (2023). Perceptions and acceptance of Artificial Intelligence: a multi-dimensional study. *Social Sciences* 12: 502. https://doi.org/10.3390/socsci12090502

Gerlich, M. (2025). AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies* 15, 1: 6. https://doi.org/10.3390/soc15010006

Ghandour, D. A. M. (2024). Navigating the impact of AI integration in higher education: ethical frontiers. U: N. H. Al Harrasi i M. S. El Din (ur.) *Utilizing AI for Assessment, Grading, and Feedback in Higher Education.* (str. 212–233). Hershey PA: IGI Global.

Ghosh, A. (2023). How can artificial intelligence help scientists? A (non-exhaustive) overview. U: A. Nolan (ur.) *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research* (str. 103–112). Paris: OECD Publishing. https://doi.org/10.1787/a8d820bd-en

Gleick, J. (2011). *The information: a history, a theory, a flood*. New York: Pantheon Books.

Grace i sur. (2018). Grace, K.; J. Salvatier; A. Dafoe; B. Zhang i O. Evans, O. When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research* 62: 729–754. [citirano: 2025–07–02]. Dostupno na: https://jair.org/index.php/jair/article/view/11222/26431

Harari, Y. N. (2024). *Neksus: kratka povijest informacijskih mreža od kamenog doba do umjetne inteligencije*. Zagreb: Fokus komunikacije.

Hogenhout, L. i Takahashi, T. (2022). *A Future with AI: voices of global youth: final report*. New York: United Nations Office of Information and Communications Technology. [citirano: 2025–07–02]. Dostupno na: https://unite.un.org/news/future-ai-voices-global-youth-report-launched

Hornberger, M.; A. Bewersdorff i C. Nerdel (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence* 5: 100165. https://doi.org/10.1016/j.caeai.2023.100165

Isiaku i sur. (2024). Isiaku, L.; , A. F. Kwala; K. U. Sambo i H. H. Isiaku (2024). Academic evolution in the age of ChatGPT: an in-depth qualitative exploration of its

influence on research, learning and ethics in higher education. *Journal of University Teaching & Learning Practices* 21, 6: 1–25. https://doi.org/10.53761/7egat807

Khalil, M. i E. Er (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. U: P. Zaphiris i A. Ioannou (ur.). *Learning and Collaboration Technologies, Lecture Notes in Computer Science* 14040 (str. 475–487). Cham: Springer. https://doi.org/10.1007/978-3-031-34411-4_32

Khatri, B. B. i P. D. Karki (2023). Artificial Intelligence (AI) in higher education: growing academic integrity and ethical concerns. *Nepalese Journal of Development and Rural Studies* 20, 1: 1–7. https://doi.org/10.3126/njdrs.v20i01.64134

Kong, S-C.; W. M-Y. Cheung i G. Zhang (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence* 2: 100026. https://doi.org/10.1016/j.caeai.2021.100026

Kosmyna i sur. (2025). Kosmyna, N.; E. Hauptmann; Y. T. Yuan; J. Situ; X-H. Liao; A. V. Beresnitzky; I. Braunstein; P. Maes. your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. *ArXiv:2506.08872*. https://doi.org/10.48550/arXiv.2506.08872

Kralj i sur. (2024). Kralj, L.; A. Blažić; H. Valečić; S. Janeš; V. Blašković; N. Marinić; K. Slišurić i suradnici. *Umjetna inteligencija u obrazovanju: edukativni priručnik o primjeni umjetne inteligencije u učenju i poučavanju za učitelje, nastavnike i stručne suradnike u školama*. Zagreb: Agencija za elektroničke medije i UNICEF. [citirano: 2025–07–02]. Dostupno na: https://www.medijskapismenost.hr/wp-content/uploads/2024/04/Umjetna-inteligencija-u-obrazovanju.pdf

Kwon, D. (2025). Is it OK for AI to write science papers? Survey reveals deep divide. *Nature*. [citirano: 2025–07–02]. Dostupno na: https://www.nature.com/articles/d41586-025-01463-8

Leung, M. i Niazi, S. (2023). *Universities on alert over ChatGPT and other AI-assisted tools*. University World News. [citirano: 2025–07–10]. Dostupno na: https://www.universityworldnews.com/post.php?story=20230222132357841

Liang i sur. (2023). Liang, W.; Yuksekgonul, M.; Mao, Y.; Wu, E. i Zou, J. GPT detectors are biased against non-native English writers. *Patterns* 4, 7: 100779. https://doi.org/10.1016/j.patter.2023.100779

Long, D. i Magerko, B. (2020). What is AI literacy? Competencies and design considerations. U: *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (str. 1–16). New York: Association for Computing Machinery.

Luckin i sur. (2022). Luckin, R.; M. Cukurova.; C. Kent i B. du Boulay. Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence* 3: 100076. https://doi.org/10.1016/j.caeai.2022.100076

Lund i sur. (2025). Lund, B. D.; T. H. Lee; N. R. Mannuru i N. Arutla. AI and academic integrity: exploring student perceptions and implications for higher education. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-025-09613-3

Mahmud i sur. (2024). Mahmud, M. M.; S. F. Wong; M. S. Mohd A'seri; R. Ahmad; Y. Yaacob; A. Qazi; N. I. Mustamam; U. Nagasundram. Exploring academic perceptions and challenges in AI integration in higher education. U: *ICETC '24: Proceedings of the 2024 the 16th International Conference on Education Technology and Computers.* (str. 167–172). New York: Association for Computing Machinery. https://doi.org/10.1145/3702163.370241

McCulloch, W. S. i W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 5: 115–133. https://doi.org/10.1007/BF02478259

McGrath i sur. (2023). McGrath, C.; T. C. Pargman; N. Juth i P. J. Palmgren. University teachers' perceptions of responsibility and artificial intelligence in higher education: an experimental philosophical study. *Computers and Education: Artificial Intelligence* 4: 100139. https://doi.org/10.1016/j.caeai.2023.100139

Miao i sur. (2021). Miao, F.; W. Holmes; R. Huang i H. Zhang. *AI and education: guidance for policy-makers*. Paris: UNESCO. [citirano: 2025–07–02]. Dostupno na: https://unesdoc.unesco.org/ark:/48223/pf0000376709

Miao, F. i W. Holmes (2023). *Guidance for Generative AI in education and research*. Paris: UNESCO. [citirano: 2025–07–02]. Dostupno na: https://unesdoc.unesco.org/ark:/48223/pf0000386693

Mrnjaus, K.; S. Vrcelj i S. Kušić (2023). Umjetna inteligencija i obrazovanje: suparnici ili saveznici?. *Jahr: Europski časopis za bioetiku* 14/2, 28: 429–445. https://doi.org/10.21860/j.14.2.9

Msambwa, M.; Z. Wen i K. Daniel (2025). The Impact of AI on the personal and collaborative learning environments in higher education. *European Journal of Education, Research, Development and Policy* 60, 1: e12909. https://doi.org/10.1111/ejed.12909

Muthanna, A. (2023). *Hong Kong University to scrap ChatGPT ban for students from September*. The HK HUB. [citirano: 2025–07–10]. Dostupno na: https://thehkhub.com/hong-kong-university-to-scrap-chatgpt-ban-for-students-from-september-2023/

Ndungu, M. (2024). Integrating basic artificial intelligence literacy with media and information literacy in higher education. *Journal of Information Literacy* 18, 2: 122–139. https://doi.org/10.11645/18.2.641

Neves i sur. (2024). Neves, J. R.; S. M.de Oliveira; M. T. Oliveira; A. E. V. Neves i G. B. Torsoni. Technological innovations in education: a scoping review on the impact of AI on academic integrity. *Observatório de la Economía Latinoamericana* 22, 7: e5803. https://doi.org/10.55905/oelv22n7-153

Ng i sur. (2021). Ng, D. T. K.; J. K. L. Leung; S. K. W. Chu i M. S. Qiao. Conceptualizing AI literacy: an exploratory review. *Computers and Education: Artificial Intelligence* 2: 100041. https://doi.org/10.1016/j.caeai.2021.100041

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. Paris: OECD Publishing. [citirano: 2025–07–02]. Dostupno na: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Olędzka i sur. (2024). Olędzka, M.; M. B.Carace; S. de Oliveira Tomaz; B. Pan; P. Jiang. AI as a teaching assistant: an innovative approach to education through customized model answer generation and guided practice. *Studia Edukacyjne* 74: 67–79. https://doi.org/10.14746/se.2024.74.4

Okonkwo, C. W. i A. Ade-Ibijola (2021). Chatbots applications in education: a systematic review. *Computers and Education: Artificial Intelligence* 2: 100033. https://doi.org/10.1016/j.caeai.2021.100033

Ouyang, F. i P. Jiao (2021). Artificial intelligence in education: the three paradigms. *Computers and Education: Artificial Intelligence* 2: 100020. https://doi.org/10.1016/j.caeai.2021.100020

Oyetola, S. O. ; B. D. Oladokun i K. Dogara (2024). LIS educators' perception towards the adoption of AI tools in Nigerian library schools. *Metaverse Basic and Applied Research* 3: 65. https://doi.org/10.56294/mr202465

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice* 20*,* 2: 1–26. https://doi.org/10.53761/1.20.02.07

Perkins i sur. (2024). Perkins, M.; L. Furze; J. Roe i J. MacVaugh. The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of Generative AI in educational assessment. *Journal of University Teaching and Learning Practice* 21, 6: 1–18. https://doi.org/10.53761/q3azde36

Rainie, L. i J. Anderson (2024). *Experts Imagine the Impact of Artificial Intelligence by 2040*. North Carolina: Imagining the Digital Future Center. [citirano: 2025–07–02]. Dostupno na: https://imaginingthedigitalfuture.org/wp-content/uploads/2024/02/AI2040-FINAL-White-Paper-2-2.29.24.pdf

Rasul i sur. (2024). Rasul, T.; S. Nair; D. Kalendra; M. Balaji; F. Santini; W. Ladeira; R. Rather i suradnici. Enhancing academic integrity among students in GenAI era: a holistic framework. *The International Journal of Management Education* 22, 3: 101041. https://doi.org/10.1016/j.ijme.2024.101041

Rumelhart, D.; G. Hinton i R. Williams (1986). Learning representations by back-propagating errors. *Nature* 323: 533–536. https://doi.org/10.1038/323533a0

Sabzalieva, E. i A. Valentini (2023). *ChatGPT and Artificial Intelligence in higher education: quick start guide*. Paris: UNESCO. [citirano: 2025–07–02]. Dostupno na: https://unesdoc.unesco.org/ark:/48223/pf0000385146

Schmarzo, B. (2023). *AI & data literacy: empowering citizens of data science*. Birmingham, Dubai.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 3: 417–457. https://doi.org/10.1017/S0140525X00005756

Shen, W. i Liu, Y. (2022). China's normative systems for responsible AI: from soft law to hard law. U: S. Voeneky, P. Kellmeyer, O. Mueller i W. Burgard (ur.). *The Cambridge Handbook of Responsible Artificial Intelligence interdisciplinary perspectives*. (str. 150–166). Cambridge: Cambridge University Press.

Shneiderman, B. (2022). *Human-centered AI*. Oxford: Oxford University Press.

Song, N. (2024). Higher education crisis: academic misconduct with Generative AI. *Journal of Contingencies and Crisis Management* 31, 1: e12532. https://doi.org/10.1111/1468-5973.12532

Southworth i sur. (2023). Southworth, J.; K. Migliaccio; J. Glover; J. N. Glover; D. Reed; C. McCarty; J. Brendemuhl i A. Thomas, A. Developing a model for AI Across the curriculum: transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence* 4: 100127. https://doi.org/10.1016/j.caeai.2023.100127

Teachflow. (2023). *The evolution of AI in education: past, present, and future*. Teachflow. [citirano: 2025–07–02]. Dostupno na: https://teachflow.ai/the-evolution-of-ai-in-education-past-present-and-future/

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 236: 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vrbanus, S. (2023.). ChatGPT blokiran u Italiji zbog zabrinutosti za privatnost korisnika. *Bug*. [citirano: 2025–07–11]. Dostupno na: https://www.bug.hr/propisi/chatgpt-blokiran-u-italiji-zbog-zabrinutosti-za-privatnost-korisnika-32636

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education* 21, 1: 1 –29. https://doi.org/10.1186/s41239-024-00448-3

Wangsa i sur. (2024). Wangsa, K.; S. Karim; E. Gide i M. Elkhodr (2024). A Systematic review and comprehensive analysis of pioneering AI chatbot models from education to healthcare: ChatGPT, Bard, Llama, Ernie and Grok. *Future Internet* 16: 219. . https://doi.org/10.3390/fi16070219

Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. Cambridge, Massachusetts: MIT Press.

Wilson, A. i C. Burleigh (2025). Research integrity in the era of Generative Artificial Intelligence. *Journal of Educational Research and Practice* 15, 1: 9. https://scholarworks.waldenu.edu/jerap/vol15/iss1/9/

World Economic Forum. (2024). *Shaping the future of learning: the role of AI in education 4.0: Insight Report*. Geneva: World Economic Forum. [citirano: 2025–07–02]. Dostupno na: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Learning_2024.pdf

Xiao, P.; Y. Chen i W. Bao, W. (2023). Waiting, banning, and embracing: an empirical analysis of adapting policies for Generative AI in higher sducation. *SSRN*. https://dx.doi.org/10.2139/ssrn.4458269

Yang i sur. (2021). Yang, S. J. H.; H. Ogata; T. Matsui i N. S. Chen. Human-centered artificial intelligence in education: seeing the invisible through the visible. *Computers and Education: Artificial Intelligence* 2: 100008. https://doi.org/10.1016/j.caeai.2021.100008

Yi, Y. (2021). Establishing the concept of AI literacy: focusing on competence and purpose. *Jahr: Europski časopis za bioetiku* 12/2, 24: 353–368. https://doi.org/10.21860/j.12.2.8

Zhang, K. i A. B. Aslan (2021). AI technologies for education: recent research & future directions. *Computers and Education: Artificial Intelligence* 2: 100025. https://doi.org/10.1016/j.caeai.2021.100025