asist. Anja Brelih, mag. inž. mm.
anja.brelih@fgg.uni-lj.si
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo,
Jamova cesta 2, 1000 Ljubljana

asist. dr. Aleksander Srdić, univ. dipl. inž. grad.
aleksander.srdic@fgg.uni-lj.si
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo,
Jamova cesta 2, 1000 Ljubljana

doc. dr. Jaka Dujc, univ. dipl. inž. grad.
jaka.dujc@fgg.uni-lj.si
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo,
Jamova cesta 2, 1000 Ljubljana

doc. dr. Robert Klinc, univ. dipl. inž. grad.
robert.klinc@fgg.uni-lj.si
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo,
Jamova cesta 2, 1000 Ljubljana

Znanstveni članek
UDK/UDC: 004.434:004.8:624

PREDOBDELAVA PODATKOV ZA ZAGOTAVLJANJE VARNOSTI IN ZASEBNOSTI PRI UPORABI VELIKIH JEZIKOVNIH MODELOV V GRADBENIŠTVU

DATA PREPROCESSING TO ENSURE SECURITY AND PRIVACY WHEN USING LARGE LANGUAGE MODELS IN CONSTRUCTION

Povzetek

Prispevek predstavlja izzive zagotavljanja varstva podatkov pri uporabi velikih jezikovnih modelov (VJM) v delovnih tokovih operativnega gradbeništva. Analizira, kako uspešno obstoječa orodja za prepoznavanje imenskih entitet (angl. Named Entity Recognition, NER) zaznajo in anonimizirajo občutljive informacije v tehničnih gradbenih dokumentih, zlasti v slovenskem jeziku. Opravljena je bila kvalitativna evalvacija štirih ogrodij za obdelavo naravnega jezika (SpaCy, SpaCy SLO, Flair, NLTK), ki so bila preizkušena na vzorcu petih dejanskih gradbenih dokumentov in primerjana z ročno anotiranimi referenčnimi podatki. V evalvacijo je bila vključena tudi anonimizacija z VJM, ki je občutljive podatke zakrival z uporabo regularnih izrazov. Rezultati kažejo, da je osnovna anonimizacija sicer mogoča, vendar vsa klasična ogrodja NER slabše prepoznavajo entitete, specifične za področje, kot so projektne šifre, inženirski nazivi ter strukturirani številčni podatki. Ugotovitve kažejo na potrebe po prilagojenih orodjih za predobdelavo, saj netočna anonimizacija predstavlja pravna in etična tveganja pri vključevanju VJM v regulirane panoge, kot je gradbeništvo. Prihodnje raziskave se morajo osredotočiti na gradnjo hibridnih anonimizacijskih tokov in učenje modelov na anotiranih podatkih, da bi izboljšali natančnost in skladnost v tehničnih panogah.

Ključne besede: veliki jezikovni modeli, zasebnost podatkov, prepoznavanje imenskih entitet, operativno gradbeništvo, predobdelava dokumentov

Summary

This article addresses the challenge of ensuring data privacy when using Large Language Models (LLMs) in Construction Management Workflows. It analyses how effectively existing Named Entity Recognition (NER) tools can identify and redact sensitive information in technical construction documents, particularly in the Slovenian language. A qualitative evaluation was performed using four NLP frameworks (SpaCy, SpaCy SLO, Flair, NLTK) applied to a sample of five real-world construction documents and compared with manually annotated baseline data. The evaluation also included anonymization with VJM, which masked sensitive data using regular expressions. The results show that although basic anonymization is possible, all classical NER frameworks underperform in identifying domain-specific entities such as project codes, engineering titles and structured numerical data. These findings highlight the urgent need for domain-adapted preprocessing tools, as inaccurate redaction poses legal and ethical risks when integrating LLMs into regulated domains such as construction. Future work should focus on building hybrid redaction pipelines and training custom models on annotated corpora to improve accuracy and compliance in technical domains.

Key words: large language models, data privacy, name entity recognition, construction management, document preprocessing

1 UVOD

Pospešeno uvajanje velikih jezikovnih modelov (VJM) in z umetno inteligenco (UI) podprtih agentov v sisteme vodenja projektov preoblikuje gradbeno panogo, saj prispeva k večji učinkovitosti, večji stopnji avtomatizacije in izboljšanemu odločanju. VJM predstavljajo pomemben potencial za obdelavo in interpretacijo velikih količin besedilnih podatkov [Ruan et al., 2023], kar lahko zagotovi boljši pregled nad ključnimi področji upravljanja gradbenih projektov, kot so ocenjevanje stroškov, analiza pogodb, projektna dokumentacija in optimizacija terminskih planov.

Vendar pa uporaba UI v gradbenih delovnih tokovih prinaša kritične izzive glede varnosti in zasebnosti občutljivih informacij [Deng et al., 2025]. Gradbena dokumentacija pogosto vsebuje osebne podatke, finančne informacije, lastniške načrte in zaupno komunikacijo. Nepravilno ravnanje s takimi podatki ali njihovo razkritje lahko povzroči resne pravne, regulativne in etične posledice. Posebej problematična je avtonomna narava UI-agentov skupaj z nepredvidljivim vedenjem VJM pri obdelavi občutljivih podatkov, kar dodatno povečuje pomisleke glede lastništva podatkov, sledljivosti in zaupanja ([Deng et al., 2025], [He et al., 2024]).

Reševanje teh pomislekov zahteva uvedbo robustnih tehnik predobdelave, ki zagotavljajo, da se občutljive informacije odstranijo ali anonimizirajo, preden se dokumenti vključijo v sisteme, ki temeljijo na VJM. Tehnike predobdelave, kot so prepoznavanje imenskih entitet (NER), psevdonimizacija, anonimizacija in odstranjevanje metapodatkov, so že bile preizkušene v drugih sektorjih ([Miranda et al., 2025], [Yermilov et al., 2023]), vendar je njihova uporaba pri specifični obliki in terminologiji gradbenih dokumentov še vedno premalo raziskana.

Kljub hitremu razvoju tehnologij VJM in tehnik varovanja zasebnosti se malo raziskav ukvarja z varno predobdelavo, posebej prilagojeno gradbeni dokumentaciji. Večina obstoječih raziskav se ukvarja s splošnimi tveganji zasebnosti na ravni modela UI, ne pa na ravni priprave dokumentov. Gradbeni dokumenti se od standardnih besedil razlikujejo po obliki, terminologiji in občutljivosti, zato je prilagoditev obstoječih tokov anonimizacije in zakrivanja podatkov ključen raziskovalni izziv.

Glavni cilj tega prispevka je oceniti učinkovitost tehnik predobdelave dokumentov pri odstranjevanju občutljivih podatkov pred uporabo VJM v delovnih tokovih upravljanja gradbenih projektov. Raziskava se posebej osredotoča na to, kako omogočiti varno pripravo dokumentov za zmanjšanje tveganj zasebnosti pri uporabi UI-agentov, ki temeljijo na VJM, v digitalnih okoljih gradbenih projektov.

Raziskava temelji na predhodnem delu avtorjev [Brelih et al., 2025], z dopolnitvijo, kjer uporabimo sam VJM za prepoznavanje in zakrivanje občutljivih podatkov v dokumentu. Za dosego tega cilja si raziskava zastavlja naslednje cilje:

1. preizkusiti več obstoječih orodij za NER in preveriti, kako uspešna so pri prepoznavanju in zakrivanju takih podatkov, zlasti v dokumentih v slovenskem jeziku;
2. preveriti uspešnost prepoznavanja in zakrivanja občutljivih podatkov v dokumentih z velikim jezikovnim modelom, ki uporablja regularne izraze (angl. regular expression, regex);
3. primerjati rezultate teh orodij z ročno anotiranimi primeri, da bi ocenili njihovo relativno natančnost;
4. na podlagi ugotovitev ponuditi preliminarne predloge za izboljšanje predobdelave dokumentov v kontekstu gradbenih delovnih tokov.

Ta raziskava ima več omejitev. Analiza je omejena na majhen vzorec besedilnih digitalnih dokumentov v formatu PDF, pri čemer so bili optično razpoznani ali slikovni dokumenti izključeni zaradi dodatne zahtevnosti prepoznavanja in razčlenjevanja besedila. Prispevek se osredotoča izključno na ukrepe predobdelave, pri čemer tveganja zasebnosti in varnosti, povezana s fazami učenja modela VJM, niso zajeta. Čeprav so na voljo kvantitativne metrike, kot so natančnost (angl. precision), priklic (angl. recall) in mera F1 za proces anonimizacije, v tem delu niso bile uporabljene znotraj VJM, temveč le za primerjavo orodij NER ter uporabo VJM za prepoznavanje in zakrivanje entitet z ročno anotiranimi podatki.

Preostanek prispevka je organiziran tako, da drugo poglavje prinaša pregled literature s področja gradbene informatike, UI-agentov in tehnik predobdelave za varnost podatkov. V tretjem poglavju je podan kratek pregled velikih jezikovnih modelov s poudarkom na njihovi arhitekturi in vedenjskih značilnostih, ki so pomembne za predobdelavo dokumentov. Četrto poglavje opisuje uporabljeno metodologijo, peto pa njeno uporabo na izbranem naboru gradbenih dokumentov. V šestem poglavju so predstavljene ugotovitve raziskave, sedmo poglavje vsebuje razpravo, osmo poglavje pa sklepne ugotovitve ter predloge za nadaljnje raziskave.

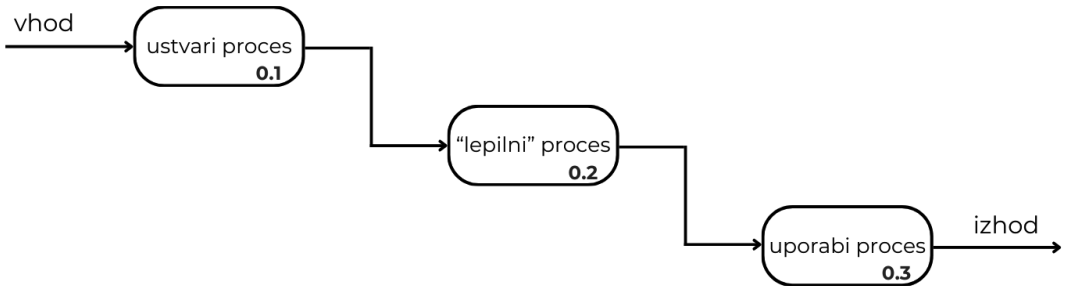
2 PREGLED LITERATURE

V poglavju so povzete ključne raziskave in koncepti, ki se nanašajo na gradbeno informatiko, razvoj od Gradbeništva 3.0 do Gradbeništva 5.0, izzive varnosti in zasebnosti pri vključevanju VJM v delovne tokove ter obstoječe tehnike za zagotavljanje varnosti podatkov in predobdelavo dokumentov.

2.1 Gradbena informatika in vodenje projektov

Gradbeništvo kot tehniška panoga izstopa po edinstvenih značilnostih, kot je izdelava enkratnih izdelkov, ki zahtevajo specifično zaporedje procesov ter sodelovanje različnih partnerjev [Turk, 2006]. Prav zaradi teh posebnosti se je razvila gradbena informatika, ki je definirana kot *»celota ved, strok in tehnologij, ki so namenjene sistematskemu obravnavanju podatkov in informacij v gradbeništvu«* [Turk, 2001]. Področje se ukvarja s procesi ustvarjanja, obdelave in posredovanja gradbenospecifičnih informacij, bodisi v obliki, razumljivi človeku, bodisi v strojno berljivi obliki. Ti procesi so tradicionalno zasnovani na strukturiranih odnosih vhod–izhod, pri čemer »lepilni« procesi povezujejo procese prek statičnih podatkovnih formatov, definiranih aplikacijskih programskih vmesnikov ali kodiranih skript (glej sliko 1).

Vodenje gradbenih projektov, ki je tradicionalno usmerjeno v obvladovanje časa, stroškov in kakovosti, se vse bolj širi tudi na okoljske in družbene vidike trajnosti [Taboada et al., 2023]. V



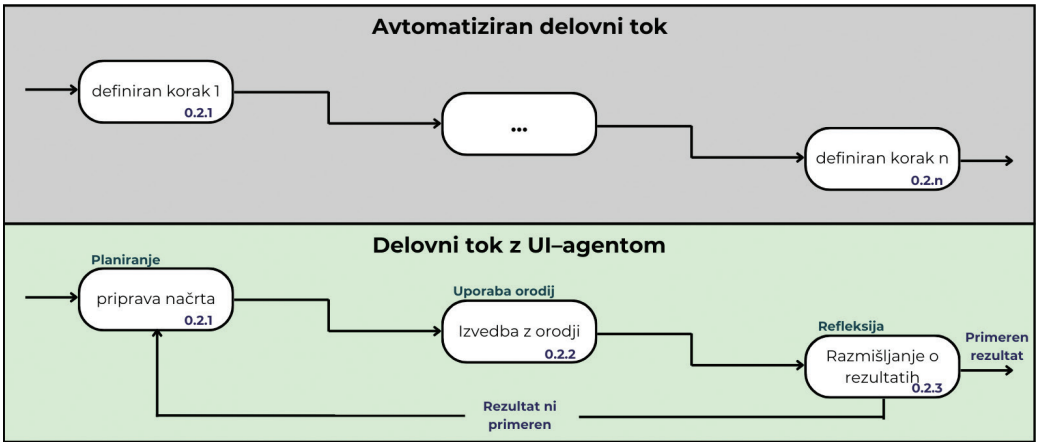
Slika 1. *Delovni tok gradbene informatike [Turk, 2006].*

tem okviru digitalna orodja, zlasti UI, pridobivajo pomen kot podporni sistemi za odločanje. Raziskave kažejo, da lahko sodobni UI-agenti, ki temeljijo na VJM, pomembno prispevajo k:

- avtomatiziranem spremljanju in poročanju, kjer agenti nenehno preverjajo skladnost z omejitvami časovnic in proračunov ter predlagajo ukrepe ob odstopanjih [Ruan et al., 2023];
- podpori odločanju, kjer z združevanjem različnih virov podatkov oblikujejo alternativne scenarije in pomagajo izbrati optimalne poti projekta [Mikhridinova et al., 2024];
- napovedovanju tveganj, kjer z analizo trenutnega stanja projekta in zgodovinskih podatkov v realnem času prepoznajo nastajajoča tveganja in predlagajo strategije za njihovo ublažitev [Taboada et al., 2023].

Takšne aplikacije so dobro prilagojene naraščajoči kompleksnosti in nepredvidljivosti sodobnih gradbenih projektov. Gradbena informatika zato zagotavlja ne le tehnološko podlago, temveč tudi metodološki okvir za vključevanje UI-agentov v dejanske gradbene delovne tokove. Medtem ko avtomatizirani delovni tokovi sledijo vnaprej določenim korakom, UI-agenti delujejo dinamično – načrtujejo, izvajajo korake z orodji in vrednotijo rezultate (glej sliko 2).

med Gradbeništvom 3.0, 4.0 in 5.0. V fazi Gradbeništva 3.0 gre predvsem za ločeno uporabo digitalnih orodij, kjer človek prek računalnika komunicira z digitalnim okoljem, ločenim od fizičnega. Gradbeništvo 4.0 je digitalna preobrazba gradbene industrije, torej povezovanje fizičnega in digitalnega okolja prek kibernetsko-fizičnih sistemov [Klinc & Turk, 2019]. S tem se ustvarja povezan ekosistem med stroji, podatki in uporabniki. Gradbeništvo 5.0, kot razširitev Industrije 5.0, se ponovno osredotoča na človeka in spodbuja razvoj inteligentnih sistemov, ki sodelujejo z ljudmi, namesto da bi jih nadomestili [Musarat et al., 2023]. To vključuje avtonomne sisteme, ki sprejemajo odločitve in ukrepe na podlagi podatkovne analitike in algoritmov strojnega učenja. Povezan ekosistem se v Gradbeništvu 5.0 še poglobi z uvedbo kibernetsko-fizično-kognitivnih sistemov, kjer tehnologije razumejo kontekst, sodelujejo z ljudmi in se prilagajajo njihovim potrebam. Tak pristop sledi načelu človek v zanki (angl. Human-in-the-Loop, HITL), ki je eno od temeljev Industrije 5.0 [Musarat et al., 2023]. UI-agenti tako postanejo kognitivni sodelavci, ki dopolnjujejo človeško odločanje [Deng et al., 2025].



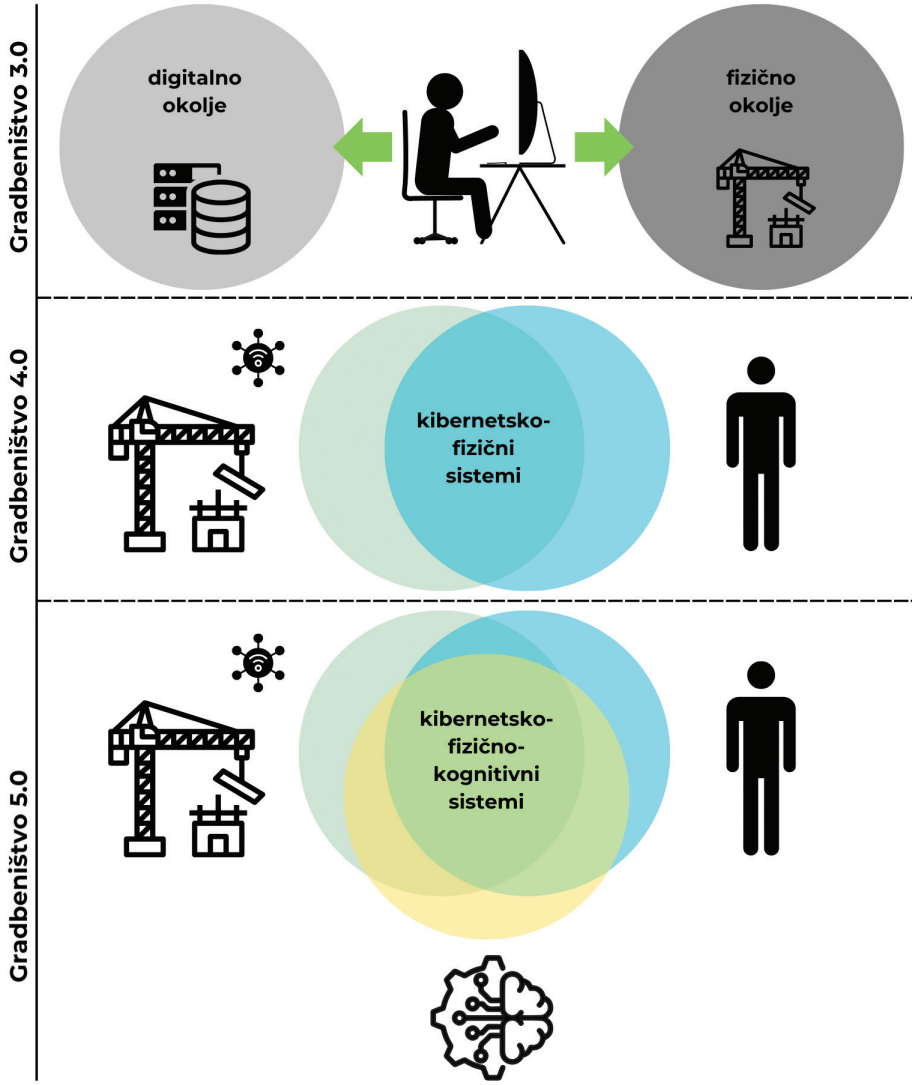
Slika 2. *Primerjava med avtomatiziranim delovnim tokom in delovnim tokom z UI-agentom [Newhauser et al., 2025].*

2.2 Od Gradbeništva 3.0 do Gradbeništva 5.0

Gradbeništvo je začelo vključevati VJM in UI-agente v delovne tokove projektnega vodenja, kar pomeni pomemben korak k paradigmi Gradbeništva 5.0. Slika 3 prikazuje razliko

2.3 Izzivi varnosti in zasebnosti pri vključevanju VJM

Uvajanje VJM je sprožilo številne pomisleke glede varnosti in občutljivih podatkov [He et al., 2024]. Gradbeni dokumenti pogosto vsebujejo finančne informacije, osebne podatke in



Slika 3. *Razlika med Gradbeništvom 3.0, 4.0 (temelji na [Klinc & Turk, 2019]) in 5.0.*

projektno-specifične lastniške vsebine. Tveganje nenamernega razkritja takih podatkov se še poveča, kadar obdelavo izvajajo modeli UI, še posebno pri uporabi javnih oblačnih storitev ali zunanjih aplikacijskih programskih vmesnikov [Deng et al., 2025].

UI-agenti ta tveganja še povečajo, saj dostopajo do več sistemov, sprejemajo samostojne odločitve in pogosto delujejo brez podrobnega nadzora [Liu et al., 2024]. Takšne zmožnosti sprožajo zapletena vprašanja glede lastništva podatkov, sledljivosti in skladnosti z zakonodajo. V reguliranih panogah, kot je gradbeništvo, veljata ohranjanje človeškega nadzora (načelo HITL) in stroga ureditev pravic dostopa za ključna mehanizma zmanjšanja tveganj [Kotsiopoulos et al., 2024].

2.4 Tehnike za varnost podatkov in predobdelavo

Za ublažitev tveganj varnosti in zasebnosti, povezanih z vključevanjem VJM v gradbene delovne tokove, je postala ključna predobdelava podatkov, ki zagotavlja varnost infromacij. To vključuje tehnike, kot so prepoznavanje imenskih entitet (NER), anonimizacija, psevdonimizacija in avtomatizirano zakrivanje podatkov, s katerimi se občutljive informacije prepoznajo in zakrijejo, še preden jih obdelajo VJM ([Miranda et al., 2025], [Yermilov et al., 2023]). Namen teh metod je preprečiti nenamerno razkritje občutljivih in zaupnih informacij.

Orodja, kot so SpaCy [SpaCy, 2025], Flair [Akbik, 2023] in NLTK [NLTK, 2024], ponujajo robustne zmožnosti za zaznavanje občutljivih podatkovnih vzorcev tako s statističnimi modeli NER kot z metodami, ki temeljijo na pravilih. Vendar pa te pristope pogosto spremlja potreba po ročnem preverjanju zaradi napačnih zaznav (angl. false positives) in zagotavljanja natančnosti, zlasti pri kompleksnih ali nestrukturiranih dokumentih.

Ker UI-agenti postajajo vse bolj avtonomni in integrirani v različne sisteme, s seboj prinašajo nove varnostne izzive. Deng in sodelavci [2025] izpostavljajo štiri ključne vrzeli v znanju o varnosti UI-agentov: (1) nepredvidljivost večstopenjskih uporabniških vnosov, (2) kompleksnost notranjih izvajanj, (3) spremenljivost operativnih okolij ter (4) interakcije z nezanesljivimi zunanjimi entitetami. Ti dejavniki še dodatno poudarjajo po-

men robustnih tehnik predobdelave, da UI-agenti lahko delujejo varno in učinkovito znotraj predvidenih omejitev.

3 PREGLED VELIKIH JEZIKOVNIH MODELOV

Veliki jezikovni modeli so v zadnjem času postali ključna tehnologija na področju obdelave naravnega jezika (ONJ), saj omogočajo obdelavo ogromnih količin podatkov in ustvarjanje koherentnih in kontekstualno ustreznih besedilnih odgovorov. Njihov razvoj je tesno povezan z napredkom na področju globokega učenja ter razpoložljivostjo velikih količin označenih in neoznačenih besedilnih podatkov ([Devlin et al., 2019], [Lewis et al., 2021]).

VJM temeljijo na arhitekturi transformerjev, ki so bili prvič predstavljeni pri [Vaswani et al., 2017]. Arhitektura transformerjev uporablja mehanizem samopozornosti (angl. self-attention), ki modelu omogoča, da prepozna, katerim delom besedila naj pripiše večji pomen glede na kontekst. To predstavlja velik odmik od prejšnjih pristopov, ki so temeljili na nevronskih mrežah s povratno zanko (angl. recurrent neural networks, RNN) ali konvolucijskih nevronskih mrežah (angl. convolutional neural networks, CNN), saj omogoča bistveno učinkovitejše učenje dolgoročnih odvisnosti med besedami.

VJM obdelujejo besedilo tako, da ga najprej razdelijo na manjše enote, ki so lahko posamezne besede ali znaki. Vsaka enota se pretvori v vektorsko predstavitev, ki omogoča računalniško obdelavo. S pomočjo mehanizma samopozornosti se nato oblikujejo kontekstualne povezave, kar modelu omogoča napovedovanje najverjetnejšega naslednjega elementa v zaporedju.

V fazi predtreniranja se modeli učijo na zelo obsežnih zbirkah besedil, kot so spletni viri, knjige in programska koda, praviloma brez eksplicitnih oznak. Takšno učenje omogoča modelom oblikovanje statističnega razumevanja jezikovne strukture, slovnice, dejstev in vzorcev. Po tej fazi lahko sledi dodatno prilagajanje na bolj specifične naloge ali podatkovne zbirke (angl. fine-tuning), v nekaterih primerih pa tudi učenje s povratnimi informacijami človeških uporabnikov (angl. Reinforcement Learning from Human Feedback, RLHF), ki izboljša usklajenost modela s človeškimi pričakovanji in etičnimi smernicami.

3.1 Temeljne značilnosti modelov

Ena najbolj značilnih lastnosti VJM je njihova sposobnost izvajanja zelo različnih jezikovnih nalog brez posebnega dodatnega učenja za vsako nalogo. Ker so bili usposobljeni na raznolikih in obsežnih zbirkah podatkov, lahko prepoznavajo vzorce in strukture na številnih področjih. Posledično lahko tvorijo koherentno besedilo, odgovarjajo na vprašanja, povzemajo dokumente, prevajajo jezike in celo izvajajo osnovna sklepanja z minimalnimi primeri (angl. few-shot learning) ali celo brez njih (angl. zero-shot learning) [Brown et al., 2020].

Ta sposobnost se z rastjo modelov še okrepi. Raziskave so pokazale, da se z večanjem števila parametrov in količine podatkov pojavijo tudi nove, prej odsotne sposobnosti, kot so aritmetično sklepanje, generiranje programske kode, večstopenjsko sklepanje in reševanje problemov [Wei et al., 2022]. Čeprav te sposobnosti niso popolnoma zanesljive, kažejo, da VJM presegajo zgolj statistično posnemanje jezika.

Pomembna značilnost VJM je tudi sposobnost sledenja navodilom, kar pomeni, da lahko naloge izvajajo na podlagi interpretacije vhodnih navodil ali primerov, vključenih v besedilo. To je vodilo k razvoju tehnik inženiringa pozivov (angl. prompt engineering), kjer se uspešnost modela optimizira z oblikovanjem jasnih in ustreznih navodil. S tem postanejo VJM izredno prilagodljivi, a hkrati občutljivi na obliko in jasnost uporabniških pozivov.

3.2 Splošno vedenje in omejitve

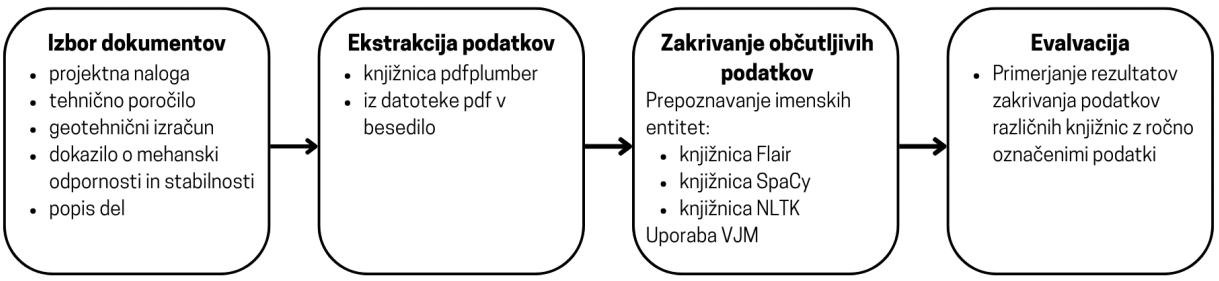
Kljub širokemu naboru zmožnosti imajo VJM tudi pomembne omejitve. Njihova učinkovitost je v veliki meri odvisna od podatkov, na katerih so bili usposobljeni, kar pomeni, da pri tehnično zahtevnih besedilih, kot so gradbeni dokumenti, lahko prihaja do netočnih rezultatov.

Pomembna omejitev so t. i. halucinacije, ko modeli ustvarijo navidezno pravilne, a dejansko napačne informacije. To predstavlja tveganje v reguliranih panogah, kjer sta natančnost in skladnost ključnega pomena. Poleg tega so VJM pogosto obravnavani kot »črne skrinjice«, saj je težko razložiti razloge za določene izhode, kar zmanjšuje transparentnost in zaupanje uporabnikov.

Te omejitve poudarjajo pomen skrbne predobdelave podatkov, zlasti v gradbeništvu, kjer je treba ustrezno obravnavati obliko in vsebino dokumentov, strokovno terminologijo ter vidike varnosti in zasebnosti podatkov.

4 METODOLOGIJA

Raziskava temelji na strukturirani kvalitativni analizi, katere namen je bil ovrednotiti učinkovitost tehnik predobdelave za zagotavljanje varnosti in zasebnosti podatkov v gradbeni dokumentaciji pred vključevanjem v VJM. V ta namen je bil zasnovan metodološki okvir, prikazan na sliki 4, ki je bil implementiran s programskim jezikom Python.



Slika 4. *Načrt metodologije zakrivanja občutljivih podatkov.*

4.1 Izbor dokumentov

Za testiranje je bil uporabljen reprezentativen nabor gradbenih dokumentov. Podatkovna zbirka je obsegala pet vrst dokumentov, ki so uporabljeni pri vodenju gradbenih projektov: (1) projektna naloga, (2) tehnično poročilo, (3) geotehnični izračun, (4) dokazilo o mehanski odpornosti in stabilnosti ter (5) popis del. Vsi izbrani dokumenti so bili v slovenskem jeziku.

4.2 Ekstrakcija podatkov

Besedilni podatki so bili iz datotek PDF pridobljeni z avtomatiziranimi orodji. Uporabljena je bila knjižnica pdfplumber [pdfplumber, 2025], ki omogoča zanesljivo razčlenjevanje besedilnih datotek PDF ter v določeni meri ohranja postavitev in obliko dokumenta. Rezultat je niz, ki vsebuje celotno besedilo dokumenta, razdeljeno po straneh. Tak izhod se nato posreduje ogrodjem za prepoznavo imenskih entitet (NER) za nadaljnje zaznavanje in zakrivanje.

4.3 Zakrivanje občutljivih podatkov

Za oceno učinkovitosti avtomatiziranih tehnik zakrivanja v gradbenih dokumentih smo implementirali in primerjali metode, ki temeljijo na NER, in sicer s tremi različnimi ogrodji za obdelavo naravnega jezika: Flair [Akbik, 2023], SpaCy [spaCy, 2025] in NLTK [NLTK, 2025]. Cilj je bil prepoznati in zakriti občutljive entitete, kot so imena oseb in organizacij ter lokacije, še preden bi dokumente obdelovali naprej.

Ogrodje Flair je bilo uporabljeno z večjezičnim modelom za označevanje zaporedij. Ta model, zasnovan na arhitekturi transformerjev, podpira NER v več jezikih ter omogoča označevanje na ravni segmentov z entitetami tipa PER (oseba), ORG (organizacija), LOC (lokacija) in MISC (razno).

Ogrodje SpaCy je bilo uporabljeno z večjezičnim modelom in z modelom za slovenščino. Slovenski model je treniran na slovenskih besedilnih korpusih, zato ponuja prepoznavanje entitet, prilagojeno slovenskemu jeziku, vključno z vrstami PER (oseba), ORG (organizacija), LOC (lokacija) in MISC (razno). Večjezični model omogoča širše jezikovno pokritje, a je manj optimiziran za slovensko skladnjo in besedišče.

Ogrodje NLTK uporablja pristop, ki temelji na pravilih. Prepoznava entitete tipa PERSON (oseba), ORGANIZATION (organizacija) in GPE (geopolitična entiteta). Omogoča preglednost in razložljivost, kar je uporabno pri osnovnem zaznavanju entitet v manjših zbirkah.

Poleg teh pristopov smo v evalvacijo vključili tudi anonimizacijo s pomočjo VJM, konkretno z modelom ChatGPT 5.1. Ta pristop ne temelji na klasičnem označevanju zaporedij, temveč na inženiringu pozivov (angl. prompt engineering), pri katerem model na podlagi besedilnega navodila samostojno prepozna in nadomesti občutljive entitete. V našem primeru je ChatGPT z uporabo regularnih izrazov generiral anonimizirano različico dokumenta ter zaznane entitete nadomestil z ustreznimi oznakami (PER, ORG, LOC, MISC).

4.4 Evalvacija

Kot je opisano zgoraj, so bile na izbranih dokumentih uporabljene štiri različne konfiguracije NER ter VJM z uporabo regularnih izrazov. Nato smo iste dokumente ročno anotirali, da smo ustvarili referenčno zbirko imenskih entitet. To je omogočilo vrednotenje rezultatov posameznega ogrodja glede na število in tipe prepoznanih entitet ter primerjavo razlik anotacij.

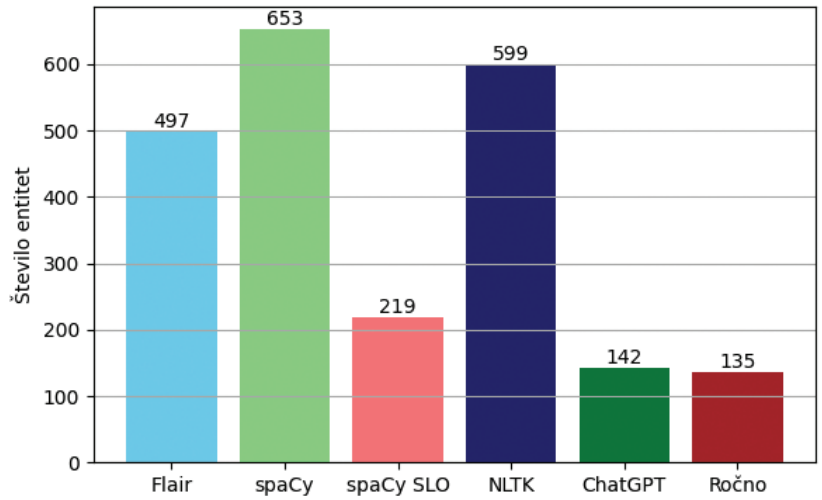
5 REZULTATI

5.1 Pregled podatkovne zbirke

Evalvacija je temeljila na petih gradbenih dokumentih v slovenščini, ki skupaj obsegajo 6631 besed. Dokumenti so vključevali več vrst občutljivih podatkov, med drugim imena oseb in organizacij, lokacije in tehnične identifikatorje.

5.2 Zaznavanje entitet

Skupno število zakritih entitet po posamezni metodi je prikazano na sliki 5. Opazno je, da je SpaCy (večjezični model) zaznal največ entitet (653), medtem ko je ročno anotiran dokument vseboval le 135 entitet. Kot prikazuje slika 5, večjezični modeli zaznajo bistveno več entitet kot referenčna zbirka, kar nakazuje na pretirano označevanje in posledično zmanj-



Slika 5. *Število entitet, zakritih z vsako metodo NER, v primerjavi z ročno anotirano referenčno zbirko (135 entitet).*

šano natančnost. Večjezični modeli so zakrivali dele besedila, ki niso občutljivi (npr. Flair je besedo poročilo označil kot organizacijo in jo zakril). V nasprotju s temi pristopi je ChatGPT zakril zgolj 142 entitet, kar je najbližje ročno anotirani referenčni zbirki, in s tem kaže bistveno manj pretiranega označevanja.

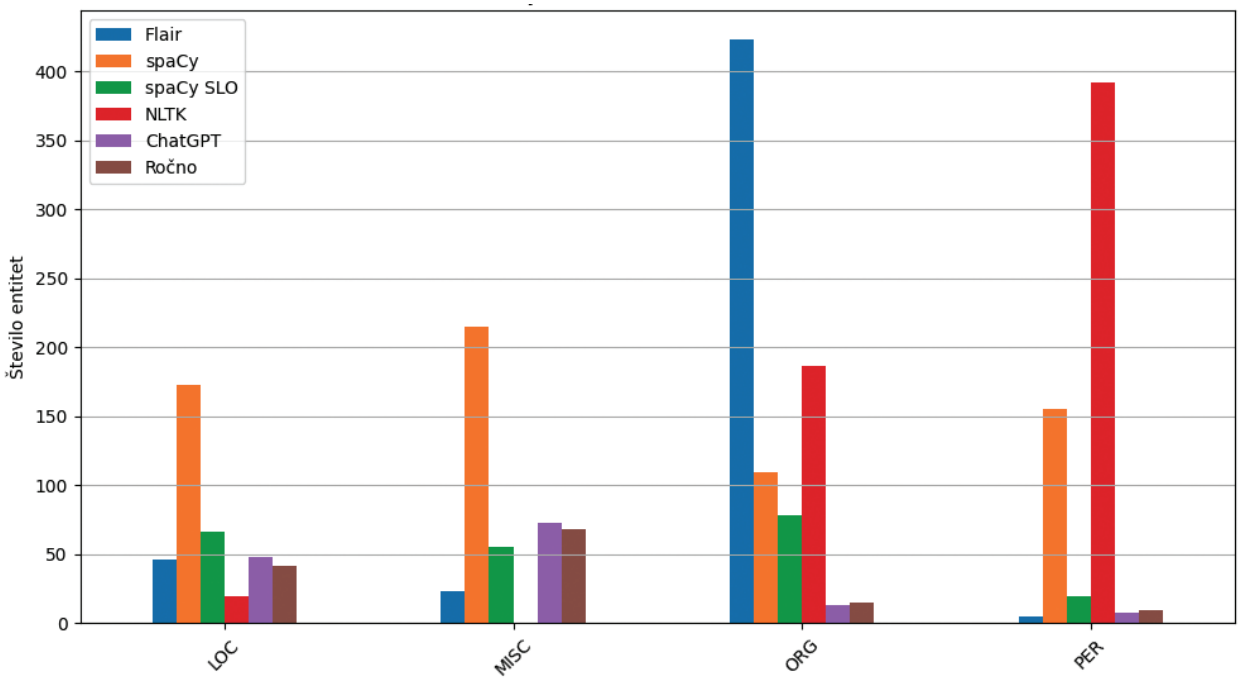
5.3 Porazdelitev tipov entitet

Število entitet je bilo nadalje razčlenjeno po tipih: PER, ORG, LOC in MISC ter normalizirano med metodami. Kot prikazuje slika 6, ogrodja izkazujejo različne pristranskosti. Ogrodje Flair pogosto pretirano označuje organizacije, slovenski model v ogrodju SpaCy deluje bolj zadržano in je skupaj označil in zakril 219 entitet, od tega 78 kot organizacije. Ogrodje NLTK kaže močno pristranskost k zaznavanju oseb in organizacij. Po drugi strani pa ChatGPT kaže bistveno bolj uravnoteženo porazdelitev entitet, kar nakazuje, da model ne kaže izrazitih pristranskosti do posameznih tipov. Število zaznanih entitet se pri ChatGPT-ju tudi najbolje približa ročni referenci.

F1. Vsak izhod ogrodja NER je bil primerjan z ročno anotirano referenčno zbirko, kjer so bile imenske entitete označene s standardnimi tipi (PER, ORG, LOC, MISC). Ker različna ogrodja uporabljajo različne oznake entitet, so bile vse oznake pred evalvacijo normalizirane na skupno shemo.

Ker se izhodi različnih orodij pogosto razlikujejo po številu in zaporedju žetonov (npr. zaradi dodanih ali odstranjenih presledkov ali znakov), bi neposredna primerjava zaporedij BIO povzročila kaskadne napake. Zato smo uvedli poravnavo zaporedij z algoritmom SequenceMatcher [difflib, 2025], ki robustno uskladi referenčne in napovedane žetone ter omogoči bolj realistično primerjavo v besedilu.

Evalvacija je upoštevala tako prisotnost kot pravilno klasifikacijo entitet. Kot prikazuje preglednica 1, so bili rezultati pri vseh ogrodjih NER nizki. Ta stroga metrika kaznuje nepravilne tipe oznak in neporavnane napovedi. Edina delna izjema je bil slovenski model v ogrodju SpaCy, ki je pokazal skromno izboljšanje zaradi prilagoditve slovenskemu jeziku.



Slika 6. *Porazdelitev zaznanih tipov entitet (PER, ORG, LOC, MISC) po metodah NER, normalizirano za primerjavo.*

5.4 Evalvacija

Za oceno učinkovitosti avtomatiziranih tehnik zakrivanja smo izvedli primerjavo na ravni žetonov (angl. token) z uporabo formata označevanja BIO (angl. inside–outside–beginning), ki natančno meri, ali so imenske entitete pravilno zaznane na ustreznem mestu v besedilu. Ta stroga metoda odraža praktična pričakovanja pri nalogah anonimizacije, zlasti kadar se dokumenti predobdelujejo za uporabo v VJM.

Evalvacija je bila izvedena z ogrodjem seqeval [Nakayama, 2018], ki omogoča metrike na osnovi zaporedij, vključno z natančnostjo (angl. precision), priklicem (angl. recall) in mero

Ogrodje	Natančnost	Priklic	Mera F1
Flair	0,115	0,491	0,187
SpaCy	0,077	0,343	0,125
SpaCy (SLO)	**0,303**	**0,463**	**0,336**
NLTK	0,050	0,278	0,085
ChatGPT 5.1	**0,733**	**0,889**	**0,803**

Preglednica 1. *Evalvacija ogrodij NER z normaliziranimi tipi entitet.*

Najbolje se je odrezal ChatGPT 5.1, ki je bistveno prehitel ogrodja NER. Model izkazuje visoko natančnost in stabilno zaznavanje lokacij entitet, kar nakazuje večjo sposobnost razumevanja domensko specifičnega konteksta.

Analiza rezultatov razkriva pomembne slabosti uporabe klasičnih ogrodij NER. Nekateri modeli so anotirali več kot 600 entitet v dokumentu s približno 6631 besedami, kar pomeni, da je bilo skoraj 10 % vseh besed označenih kot občutljivih. To kaže na visoko verjetnost pretiranega zakrivanja in lažno pozitivnih (angl. false positive) zaznav. Zakritih je bilo veliko besed, ki v resnici niso bile imenske entitete (npr. tehnični izrazi ali pogosti samostalniki v gradbeništvu).

Poleg tega so se skoraj vsa ogrodja izkazala za nezadostna pri zaznavanju entitet, specifičnih za področje, kot so:

- identifikatorji dokumentov in projektne kode,
- strokovni nazivi in certifikati inženirjev (npr. številke IZS),
- strukturirani številčni podatki, kot so telefonske številke, e-naslovi in prostorske koordinate.

Tudi slovenski model ogrodja SpaCy, čeprav bolje prilagojen jeziku, teh podatkov ni zaznaval zanesljivo.

6 UGOTOVITVE

6.1 Evalvacija učinkovitosti predobdelave

Evalvacija je pokazala, da imajo obstoječa orodja za prepoznavanje imenskih entitet velike težave pri obvladovanju specifične kompleksnosti gradbene dokumentacije. Tudi slovenski model ogrodja SpaCy, čeprav treniran na slovenskih besedilih, ni bil sposoben zanesljivo zaznati nekaterih ključnih tipov informacij, kot so projektne kode, tehnične kratice, strokovni nazivi inženirjev ter strukturirani številčni identifikatorji.

Slovenski model v ogrodju SpaCy se je izkazal za najzanesljivejšega med klasičnimi pristopi NER, vendar nobeno od teh ogrodij ni doseglo zadostne natančnosti za učinkovito anonimizacijo besedil na področju gradbeništva.

Najboljše rezultate je dosegel VJM ChatGPT 5.1, ki je izkazal višjo natančnost in bistveno manj pretiranega označevanja kot preostala ogrodja, kar nakazuje na večjo robustnost uporabe VJM in regularnih izrazov pri obdelavi tehnične dokumentacije.

Pomembno je tudi opozoriti, da so lahko na rezultate evalvacije vplivale določene nedoslednosti v anotacijskih smernicah. Na primer pri označevanju so bile vključene tudi podjetniške pripone, kot je »d. o. o.«, medtem ko večina ogrodij teh elementov ni prepoznala kot del imenskih entitet, kar je lahko prispevalo k nižji izmerjeni učinkovitosti.

6.2 Priporočene prakse

Glavne ugotovitve raziskave so:

- jezikovno specifični modeli prinašajo določene izboljšave, vendar sami po sebi niso zadostni,
- obstoječi modeli brez prilagajanja pogosto pretirano označujejo ali pa spregledajo pomembne entitete, kar vodi do lažno pozitivnih in lažno negativnih rezultatov,
- ročna anotacija je še vedno nujna za razvoj učinkovitih in zanesljivih tokov anonimizacije na tehničnih področjih,
- uporaba VJM predstavlja obetavno smer, vendar je treba zagotoviti ustrezne varnostne prakse; modele je priporočljivo izvajati v lokalnem okolju ali pa jih uporabljati zgolj za generiranje programske kode z regularnimi izrazi, ne pa za neposredno obdelavo občutljivih dokumentov.

7 DISKUSIJA

Rezultati te raziskave poudarjajo pomen specifičnih tokov obdelave naravnega jezika pri uporabi VJM z občutljivo tehnično dokumentacijo. Gradbeno področje vključuje širok nabor specifičnih podatkovnih tipov, med drugim pravne entitete, certifikate inženirjev, identifikatorje, prostorske reference in tehnične kode, ki jih splošni modeli pogosto spregledajo.

Evalvacija je pokazala, da večina klasičnih orodij NER v takšnem okolju ni zanesljiva, predvsem zaradi pretiranega označevanja in težav pri zaznavanju strukturiranih ter večbesednih entitet.

Za bistveno izboljšanje postopkov zakrivanja podatkov je potreben pristop NER, prilagojen področju, zlasti pri dokumentih, ki so napisani v jezikih, kot je slovenščina. To vključuje:

- učenje prilagojenih modelov na anotiranih gradbenih dokumentih,
- razširitev nabora oznak entitet tako, da zajema identifikatorje, elektronske naslove in druge strukturirane podatke,
- kombinacijo metod na podlagi pravil z NER za zaznavanje vzorcev, ki jih zgolj strojno učenje težko prepozna.

Takšni hibridni tokovi bi omogočili večjo natančnost, zanesljivost in preglednost, zaradi česar bi bili primernejši za zagotavljanje skladnosti z zakonodajo in praktično uporabo v gradbenih delovnih tokovih. Kot posebej obetavna se izkaže uporaba VJM, kot je ChatGPT 5.1, ki se je izkazal za bistveno bolj natančnega in manj nagnjenega k pretiranemu označevanju. Vendar je njegova uporaba pri občutljivi dokumentaciji smiselna le v varnih, lokalnih okoljih ali kot pomoč pri generiranju programske kode za anonimizacijo, ne pa za neposredno obdelavo izvirnih dokumentov.

Predstavljeno delo ponazarja omejitve splošnih ogrodij NER pri zakrivanju občutljivih podatkov v gradbenih dokumentih. Trenutno so za zagotavljanje skladnosti s standardi varnosti in zasebnosti ob uporabi VJM še vedno potrebni ročni pregledi.

8 SKLEP

Prispevek je ovrednotil, kako uspešno obstoječa orodja za prepoznavanje imenskih entitet obvladujejo zakrivanje občutljivih podatkov v slovenskih gradbenih dokumentih. Rezultati so pokazali, da je osnovna anonimizacija sicer mogoča, vendar trenutna ogrodja ne dosegajo ravni natančnosti, ki bi bila potrebna za zanesljivo rabo na tem področju. Razlog je ta, da pogosto pretirano označujejo in nezanesljivo zaznavajo domensko specifične entitete, kar omejuje njihovo uporabnost v tehničnih okoljih.

Za učinkovito predobdelavo tehničnih dokumentov pred vključevanjem v VJM je nujno doučiti modele ogrodij NER na jezikovno specifičnih, anotiranih gradbenih besedilih. Obenem je treba vključiti metode, ki temeljijo na pravilih, za zaznavanje strukturiranih elementov, kot so kode in kontaktni podatki, ter vzpostaviti hibridne tokove zakrivanja, ki združujejo jezikoslovno znanje s statističnimi modeli. Rezultati kažejo, da lahko veliki jezikovni modeli ob ustreznih varnostnih ome-

jitvah in uporabi v lokalnem okolju pomembno prispevajo k natančnejšemu in kontekstno bolj občutljivemu zakrivanju.

Prihodnje delo bo usmerjeno v zbiranje anotiranih podatkov iz dejanskih inženirskih projektov ter testiranje učinkovitosti modelov na raznolikejših tipih dokumentov. Cilj je razviti robusten in področno specifičen okvir za anonimizacijo, ki bo zagotavljal varnost in zasebnost, hkrati pa ohranjal tehnično natančnost gradbene dokumentacije.

9 ZAHVALA

Raziskavo je podprla Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije v okviru financiranja mladih raziskovalcev in raziskovalnega programa E-Gradbeništvo (P2-0210).

10 LITERATURA

Akbik, A., flair, https://flairnlp.github.io/, 2023.

Brelih, A., Srdic, A., Klinc, R. , Evaluation of privacy and security measures when using LLMs for construction management, Proceedings of the 14th Creative Construction Conference (CCC 2025), International Association for Automation and Robotics in Construction (IAARC), https://doi.org/10.22260/ccc2025/0068, 2025.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M ., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., Language Models are Few-Shot Learners, Advances in Neural Information Processing Systems, 33, 1877–1901, https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html, 2020.

Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., Xiang, Y., AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways, ACM Computing Surveys, 57(7), 1–36, https://doi.org/10.1145/3716628, 2025.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv, https://doi.org/10.48550/arXiv.1810.04805, 2019.

difflib, https://docs.python.org/3/library/difflib.html, Python, 2025.

He, Y., Wang, E., Rong, Y., Cheng, Z., Chen, H., Security of AI Agents, arXiv. https://doi.org/10.48550/arXiv.2406.08689, 2024.

Klinc, R., Turk, Ž., Construction 4.0 – Digital Transformation of One of the Oldest Industries, Economic and Business Review, 21(3), https://doi.org/10.15458/ebr.92, 2019.

Kotsiopoulos, T., Papakostas, G., Vafeiadis, T., Dimitriadis, V., Nizamis, A., Bolzoni, A., Bellinati, D., Ioannidis, D., Votis, K., Tzovaras, D., Sarigiannidis, P., Revolutionizing defect recognition in hard metal industry through AI explainability, human-in-the-loop approaches and cognitive mechanisms, Expert Systems with Applications, 255, 124839, https://doi.org/10.1016/j.eswa.2024.124839, 2024.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D., Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, arXiv, https://doi.org/10.48550/arXiv.2005.11401, 2021.

Liu, Z., Yao, W., Zhang, J., Yang, L., Liu, Z., Tan, J., Choubey, P. K., Lan, T., Wu, J., Wang, H., Heinecke, S., Xiong, C., Savarese, S., AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System (Version 1), arXiv, https://doi.org/10.48550/ARXIV.2402.15538, 2024.

Mikhridinova, N., Yurdakul, F. A., Wolff, C., Using Large Language Models for Project Staffing: Evaluation of GPT-Based Mapping of Teams to Projects, Proceedings of the 12th IPMA Research Conference "Project Management in the Age of Artificial Intelligence", 1–19, https://doi.org/10.56889/cykr4175, 2024.

Miranda, M., Ruzzetti, E. S., Santilli, A., Zanzotto, F. M., Bratières, S., Rodolà, E., Preserving Privacy in Large Language Models: A Survey on Current Threats and Solutions, arXiv, https://doi.org/10.48550/arXiv.2408.05212, 2025.

Musarat, M. A., Irfan, M., Alaloul, W. S., Maqsoom, A., Ghufran, M., A Review on the Way Forward in Construction through Industrial Revolution 5.0, Sustainability, 15(18), 13862, https://doi.org/10.3390/su151813862, 2023.

Nakayama, H., seqeval: A Python framework for sequence labeling evaluation, https://github.com/chakki-works/seqeval, 2018.

Newhauser, M., Monigatti, L., Yadav, P., Çelik, T., What Are Agentic Workflows? Patterns, Use Cases, Examples, and More, Weaviate, https://weaviate.io/blog/what-are-agentic-workflows, 2025.

NLTK , https://www.nltk.org/, Natural Language Toolkit, 2024.

pdfplumber, https://pypi.org/project/pdfplumber/, Python Package Index, 2025.

Ruan, J., Chen, Y., Zhang, B., Xu, Z., Bao, T., Du, G., Shi, S., Mao, H., Li, Z., Zeng, X., Zhao, R., TPTU: Large Language Model-based AI Agents for Task Planning and Tool Usage (Version 3), arXiv, https://doi.org/10.48550/ARXIV.2308.03427, 2023.

spaCy, Industrial-strength Natural Language Processing in Python, https://spacy.io/, 2025.

Taboada, I., Daneshpajouh, A., Toledo, N., De Vass, T., Artificial Intelligence Enabled Project Management: A Systematic Literature Review, Applied Sciences, 13(8), 5014, https://doi.org/10.3390/app13085014, 2023.

Turk, Ž., Gradbena informatika—Definicija področja, teme, problem , Gradbena Informatika 2001, Ob 30 Letnici Inštituta IKPIR, Zbornik Seminarja, 37–44, 2001.

Turk, Ž., Construction informatics: Definition and ontology. Advanced Engineering Informatics, 20(2), 187–199. https://doi.org/10.1016/j.aei.2005.10.002, 2006.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., Attention Is All You Need, arXiv, https://doi.org/10.48550/arXiv.1706.03762, 2017.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., Fedus, W., Emergent Abilities of Large Language Models, arXiv, https://doi.org/10.48550/arXiv.2206.07682, 2022.

Yermilov, O., Raheja, V., Chernodub, A., Privacy- and Utility-Preserving NLP with Anonymized data: A case study of Pseudonymization, Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023), 232–241, Association for Computational Linguistics, https://doi.org/10.18653/v1/2023.trustnlp-1.20, 2023.