KOFETKE.

44 / 50

VODNIKI O AI

Modeli kitajskega izvora brez navijaškega prepira

Kako ločeno oceniti zmožnosti, licence, gostovanje in ravnanje s podatki?



IZBIRA MODELOV

Primerjalni okvir za točno imenovane izdaje modelov in ponudnike storitev.

SLOVENSKA IZDAJA

september 2026
Praktična uporaba AI

Viri preverjeni 25. septembra 2026.

Izvor modela je slab nadomestek za pregled tega, kar bo podjetje dejansko uporabljalo. Zavrnitev izdaje zato, ker je kitajska, in njeno sprejetje zato, ker tekmuje z ameriškim ponudnikom, vsebujeta isto praktično napako: nobena odločitev ne potrdi dovoljenja za uporabo uteži, sprejemljivega ravnanja s podatki strank ali zanesljivosti pri konkretni nalogi.

Izvor lahko odpre pomembna vprašanja o razvijalcu, pogodbeni stranki, pravni pristojnosti in dobavni verigi. Sam pa nanje ne odgovori. Članek 09 je predstavil možnosti postavitve, članek 42 pa podrobneje obravnava licence uteži. Tukaj želimo identiteto izdaje, pravna dovoljenja, delovanje storitve in kakovost pri nalogi ohraniti v ločenih stolpcih ter sprejeti odločitev, katere podlage je mogoče preveriti.

Pred ocenjevanjem družine določite izdajo

»Qwen« in »DeepSeek« sta imeni družin in organizacij, ne ponovljivih preizkusnih konfiguracij. Dva konkretna primera sta **Qwen/Qwen3-8B** in **deepseek-ai/DeepSeek-R1**. Prva modelna kartica opredeljuje Qwen3-8B in navaja 8,2 milijarde parametrov. Druga pri izvirnem polnem DeepSeek-R1 navaja skupno 671 milijard parametrov in 37 milijard aktiviranih parametrov.**[1][3]** To so podatki izdajateljev, ne meritve na strojni opremi ali razvrstitev uporabnosti.

Polne izdaje R1 ne zamenjajte neopazno z destilirano različico ali poznejšo izdajo, pri tem pa ohranite iste oznake v primerjavi. Kartica R1 ločeno navaja destilirane modele, ki temeljijo na

modelih Qwen in Llama, ter opredeljuje njihove izvorne družine. [3] Tudi kvantizacija, ki jo pripravi skupnost, je ločena datoteka, ki jo je treba evidentirati: ohranite ime njenega izdajatelja, izvorno revizijo, metodo pretvorbe in zgoščeno vrednost datoteke. Ime v meniju klepetalne aplikacije ne potrjuje, katera od teh različic je ustvarila odgovor.

Primerjalni zapis naj zato vsebuje identifikator izdaje, revizijo repozitorija, datoteko uteži ali ponudnikovo različico modela, tokenizator, izvajalno okolje in pomembne nastavitve. Kadar ponudnik razkrije samo oznako, za katero se model sčasoma spreminja, zabeležite to omejitev in datum preizkusa. Takšno vrednotenje storitve je lahko uporabno, ni pa ponovljiva primerjava nespremenjenih uteži. Pri nobeni od spodaj obravnavanih trenutnih gostovanih storitev ne predpostavljamo, da izvaja katero od teh točnih izdaj zgolj zato, ker je na njeni spletni strani sorodna znamka.

Preberite dovoljenja izdaje, ne sklepajte iz njenega ugleda

Kartica Qwen3-8B napotuje na Apache License 2.0, besedilo te licence pa je v njegovem repozitoriju.[1][2]

Kartica polnega DeepSeek-R1 navaja, da sta repozitorij kode in uteži pod licenco MIT; pripadajoča licenčna datoteka vsebuje veljavno obvestilo in pogoje.[3][4] Dokumente berite skupaj. Kartica določa izdajo in navedeni obseg licence, licenca pa obveznosti, ki jih kratka oznaka »odprt« zamolči.

Točen primer	Ugotovljena licenca	Obveznosti, pomembne pri pregledu redistribucije
Qwen/Qwen3-8B	Apache License 2.0	Priložiti licenco, označiti spremenjene datoteke, ohraniti ustrezna obvestila in morebitno priloženo datoteko NOTICE obravnavati po zahtevah licence.[2]
Izvirni **deepseek-ai/DeepSeek-R1**	MIT; obseg, ki vključuje kodo in uteži, je naveden v kartici	V kopije ali bistvene dele vključiti obvestilo o avtorskih pravicah in dovoljenju.[3][4]

Ta dovoljenja so obsežna, vendar ne jamčijo, da je vsaka predvidena uporaba zakonita. Apache 2.0 vsebuje opredeljeno patentno dovoljenje in pogoj za njegovo prenehanje, omejuje dovoljenje za uporabo znamk ter izključuje jamstva, tudi glede nekršenja pravic. Tudi MIT programsko opremo daje na voljo v obstoječem stanju in izključuje jamstva.[2][4] Nobeden od dokumentov ni pogodba o zasebnosti za račun gostovane klepetalne storitve. Prav tako vam ne podeli pravic do zaupnih dokumentov stranke zgolj zato, ker lahko model zaženete.

Obseg je pomemben tudi znotraj družine. Kartica R1 pri določenih destiliranih različicah izrecno navaja izvorne licence Qwen in Llama.[3] Ta članek ne vsebuje celovitega licenčnega pregleda teh različic. Pregledovalec mora slediti dejanskemu

izvoru, ne pa oznake MIT za polni R1 pripisati vsaki datoteki z »R1« v imenu. Notranja uporaba, redistribucija spremenjenega modela in gostovani izdelek lahko pomenijo različne obveznosti; pred pravno presojo opišite načrtovano dejavnost.

Gostovana storitev doda drugo pogodbo

Pogoji uporabe DeepSeek, pridobljeni za ta članek, imajo datum 27. marec 2026. Kot upravljavca navajajo Hangzhou DeepSeek Artificial Intelligence Co., Ltd., v opis storitev vključujejo spletne strani, aplikacije, SDK-je in API-je ter določajo, da ob neskladju posebni pogoji prevladajo nad splošnimi.[6] To je izhodišče za pregled pogodbe o storitvi, ne zagotovilo, da imajo vse končne točke, stranke ali regije enake pogoje.

Ločeno pridobljeni pravilnik o zasebnosti ima datum 10. februar 2026. Velja za osebne podatke, obdelane v storitvah, ki nanj povezujejo ali se nanj sklicujejo. Izrecno izključuje pravila obdelave osebnih podatkov, zbranih od končnih uporabnikov prek aplikacij, ki so jih razvijalci zgradili s storitvami odprte platforme; ti razvijalci morajo razkriti svoje ustrezne pravilnike.[5] Ponudnik programske opreme zato ne more na vsa vprašanja o svoji aplikaciji, ki uporablja DeepSeek, odgovoriti le s povezavo na ta pravilnik. Pregledati je treba tudi njegovo lastno zbiranje, shranjevanje in nadaljnje prenose podatkov.

Pravilnik za storitve v svojem obsegu navaja, da lahko uporabniški vnosi vključujejo navodila modelu, naložene datoteke in zgodovino pogovorov. Med nameni uporabe osebnih podatkov navaja tudi učenje in izboljševanje modelov.[5] Med pravicami, ki

so lahko na voljo glede na kraj bivanja in izjeme po veljavnem pravu, navaja možnost izključitve uporabe za učenje ali optimizacijo tehnologije.[5]

Pogoji uporabe posebej obravnavajo tudi vhode in izhode. V točki 4.3 navajajo, da jih sme DeepSeek v minimalnem obsegu uporabljati za delovanje ali izboljševanje storitev in podpornih tehnologij ob navedenih pogojih šifriranja, stroge deidentifikacije in onemogočene ponovne identifikacije posameznikov. Za izključitev takšne obdelave navajajo izklop nastavitve »Improve the model for everyone«.[6] To je navedba pogodbenih pogojev, ne potrditev nastavitev posameznega računa ali ponudnikove tehnične izvedbe. Izbira glede učenja ne dokazuje, da so bili obstoječi dnevniki izbrisani.

Pravilnik o zasebnosti ločeno navaja, da storitve niso zasnovane ali namenjene obdelavi občutljivih osebnih podatkov, in uporabnikom naroča, naj jih ne posredujejo.[5] Izključitve uporabe za učenje ne obravnavajte kot izjeme od tega navodila. Pred pošiljanjem drugega zaupnega gradiva preverite pooblastilo za razkritje, veljavno pogodbo in dejanske nastavitve računa.

Pravilnik navaja, da DeepSeek osebne podatke neposredno zbira, obdeluje in hrani v Ljudski republiki Kitajski. Hramba se razlikuje glede na namen in druge dejavnike; med primeri navaja hrambo osebnih podatkov računa, vnosov in plačil, dokler račun obstaja, poleg drugih navedenih razlogov za hrambo.[5] Teh določil ne nadomestite z izmišljenim enotnim rokom hrambe. Slovensko podjetje mora upoštevati tudi ločena določila pravilnika za evropsko regijo. Samo branje objavljenega pravilnika še ne potrdi zakonitosti ureditve prenosov in obdelave v podjetju.

To so navedbe o objavljenih dokumentih imenovanega upravljavca, preverjenih na navedeni datum. Niso trditve, da uteži kitajskega izvora pošiljajo podatke na Kitajsko. Prav tako ne dokazujejo, kaj se je zgodilo pri dejanski zahtevi: za to potrebujemo dokaze o nastavitvah, pogodbeni pregled in ustrezno tehnično opazovanje. Licenca izdaje in pravilnik storitve odgovarjata na različna vprašanja, tudi kadar ju objavi isti razvijalec.

Enako skrbno preverite drugega ponudnika

Fireworks je uporaben primer za primerjavo. Njegova dokumentacija o ravnanju s podatki za navodila in ustvarjene odgovore odprtih modelov opisuje privzeto ničelno hrambo, če uporabnik izrecno ne izbere beleženja. Navaja pa tudi, da se metapodatki zahtev beležijo in da lahko predpomnjenje navodil nekatere njihove podatke v začasnem pomnilniku ohrani več minut.[8] Tako loči trajno shranjevanje vsebine od začasne obdelave in operativnih metapodatkov. »Ničelna hramba« ne pomeni, da storitev ne prejme podatkov.

Isti dokument navaja posebno izjemo: njegov Response API ima privzeto nastavitev **store=True**, ki podatke pogovora, vključno z navodili, odgovori in klici orodij, hrani do samodejnega izbrisa po 30 dneh. Nastavitev **store=False** to shranjevanje pogovora izklopi.[8] Kupec, ki prebere samo naslovno trditev, lahko izbere napačno končno točko ali nastavitve. Ta vir ne potrjuje celotne regije obdelave, podpisane pogodbe ali pogojev glede učenja za predlagano postavitev. Članek tudi ne trdi, da Fireworks trenutno ponuja katerega od zgornjih točnih modelov. Pred uvrstitvijo te

poti med kandidate preverite vsa ta polja. Ponudnik zunaj Kitajske potrebuje enak pogodbeni in tehnični pregled kot ponudnik na Kitajskem.

Lastna postavitev spremeni odgovornosti

Če organizacija objavljene uteži poganja na svojem računalniku, lahko ponudnika API-ja modela izloči iz poti obdelave navodil. S tem pa ne odstrani okoliške aplikacije. Lokalni klepetalni vmesnik lahko uporablja zunanjo storitev za vektorske vložitve, pošilja diagnostične dnevnike, sinhronizira zgodovino pogovorov ali ob težavi preklopi na oddaljeni model. To so možnosti postavitve, ki jih je treba preveriti, ne očitki obravnavanima izdajama. Narišite dejansko pot od naložene datoteke do vhoda modela, izhoda, dnevnikov in varnostnih kopij.

Pri lastni postavitvi določite, kdo upravlja računalnik in kdo lahko bere njegovo shrambo. Preverite avtentikacijo, omrežno izpostavljenost, posodabljanje, izhodne povezave, vsebino dnevnikov in delovanje brisanja. Tudi strežnik, ki ga najamete in upravljate sami, ima ponudnika infrastrukture. Preden postavitev opišete kot zasebno, predvideno mejo preverite na dejanski aplikaciji in njenih nastavitvah. Ločite trditev »nadzorujemo uteži« od trditve »vemo za vse kopije dokumentov«.

Izmišljena primerjava za slovensko kataloško ekipo

Izmišljeni veletrgovec Most iz dokumentov dobaviteljev pripravlja slovenske opise sestavnih delov za kolesa. Predlagani model lahko napiše osnutke opisov in izlušči navedene mere. Ne sme si

izmišljati združljivosti, spreminjati enot ali manjkajočega podatka nadomestiti s samozavestno trditvijo. Zaposleni pred objavo odobrijo vsak opis. Za prvo vrednotenje zadoščajo javni ali sintetični izvorni dokumenti, zato pri njem ne potrebujemo zaupnih podatkov strank.

Most Qwen3-8B uvrsti na predlagani ožji seznam za lastno postavitev, polni DeepSeek-R1 pa na pogojni seznam za gostovano uporabo. To ni trditev, da sta enako praktična na isti strojni opremi. Članek 43 pojasnjuje, zakaj pri oceni pomnilnika ne smemo zamenjati skupnega števila parametrov s številom aktivnih. Gostovana pot pride v nadaljnjo obravnavo le, če dobavitelj opredeli točno izdajo in zagotovi sprejemljive pogoje. Sicer ta kandidat ostane neustrezno opredeljen, namesto da bi ga neopazno nadomestili z novejšo različico za isto oznako.

Ekipa predlaga naslednji ločeni preizkusni nabor. Gre za načrtovane primere, ne za izmerjene uspehe ali napake katerega od modelov.

Skupina primerov	Predpostavljeno število	Kaj preveri pregledovalec
Običajni opisi s popolnimi izvornimi podatki	30	Vsebinsko zvest opis v slovenščini, navedene mere in sledljivost dokazov
Manjkajoče specifikacije	12	Manjkajoči podatki ostanejo označeni kot manjkajoči; model prosi za pojasnilo

Skupina primerov	Predpostavljeno število	Kaj preveri pregledovalec
Številske pasti in pasti z enotami	10	Vrednosti, enote in oznake modelov izdelkov ostanejo nespremenjene
Zahteve po nepotrjeni združljivosti	8	Brez izmišljene trditve, da del ustreza drugemu izdelku

Skupine skupaj vsebujejo 60 primerov. Z dvema kandidatnima izdajama in dvema izvedbama vsakega primera za vsako izdajo načrt predvideva 240 izhodov. Če za začetni pregled posameznega izhoda predpostavimo tri minute, dobimo 720 minut oziroma 12 ur. To je računska predpostavka za načrtovanje, ne izmerjena hitrost pregledovanja. Priprava nabora, razreševanje nestrinjanj in ponovni preizkusi zahtevajo dodaten čas. Nizka cena sklepanja tega dela ne bi odpravila.

Vsak primer potrebuje pričakovani rezultat ali izrecno pravilo sprejemljivosti, zapisano pred vpogledom v izhode modelov. Če manjka premer, pravilo zahteva navedbo, da premera ni, ne ujemanja z eno samo prednostno formulacijo. Pri vprašanju o združljivosti mora odgovor povedati, kaj vir podpira, in opustiti nepotrjeno trditev. Izvorne dokaze ločite od predlaganega besedila modela, da lahko pregledovalec opazi verjetne, a dodane podatke.

Primerjava naj bo poštena, ne umetno izenačena

Pri vseh kandidatih ohranite iste izvorne dokumente, vsebino naloge in dovoljene dokaze. Na majhnem razvojnem naboru določite podprto obliko navodil za posamezno izdajo, nato pa jo pred ločenim preizkusom zaklenite. Qwen3-8B dokumentira preklop med načinom z razmišljanjem in načinom brez njega.[1] Zabeležite izbrani način, omejitve izhoda in nastavitve generiranja. Enemu kandidatu ne namenjajte ponavljajočih se ročnih popravkov, medtem ko pri drugem ocenite le prvi odgovor.

Poštenost ne zahteva pretvarjanja, da so tokenizatorji ali izvajalna okolja enaki. Ohranite enakovredno nalogo, zabeležite dejansko število žetonov in izmerite čas na poti, ki jo obravnavate. Če ima en sistem več časa ali daljši dovoljeni izhod, to razliko v virih navedite ob kakovosti. Kartica DeepSeek-R1 pri primerjalnih preizkusih opisuje pogoje generiranja in vzorčenja. Njeni objavljeni rezultati so vrednotenje izdajatelja pod temi pogoji, ne rezultati za Mostov slovenski katalog.[3]

Kadar je izvedljivo, med pregledom skrijte oznake modelov. Ločeno ocenite zvestobo dejstvom, berljivost slovenščine, izpuščene zahtevane podatke, nepodprte dodatke in ustrezno zadržanost. Lep odstavek s spremenjenim premerom ne opravi naloge, četudi dobi visoko oceno sloga. Rezultate poročajte za vsako skupino primerov, ne le kot skupno povprečje. Običajni opisi ne smejo prikriti napak pri manjkajočih specifikacijah. Ohranite nestrinjanja in način njihove razrešitve, namesto da jih nadomestite z navidezno natančno enotno oceno.

Tudi varnostno vedenje je treba razlagati glede na nalogo. Če delo vključuje politično občutljivo, sporno ali škodljivo vsebino, dodajte ustrezne primere in natančno zabeležite odgovore, zavrnitve ter izpuste. O vedenju ne sklepajte iz nacionalnega izvora in zavrnitve enega navodila ne posplošujte na vse teme. Za Most je neposredno tveganje nepotrjena informacija o izdelku. Geopolitični preizkusi ne bi nadomestili preverjanja slovenskega izrazja in mer.

Napaka, zaradi katere se odločitev spremeni

Recimo, da Mostov osnutek nabavne presoje odobri gostovano pot, ker ima prenosljivi model permisivno licenco. Zaposleni nato v klepetalni vmesnik naloži neobjavljeni dokument dobavitelja, ne da bi pregledal pogoje tega vmesnika glede podatkov. V tem izmišljenem primeru podjetje pomeša različna dovoljenja: dovoljenje za kopiranje uteži razume kot dovoljenje za razkritje tujih informacij. Boljši rezultat modela na primerjalnem preizkusu ne bi nadomestil manjkajočega pooblastila ali pogodbenega pregleda.

Ustrezen odziv je ustaviti nove zaupne prenose, ugotoviti točno storitev in nastavitve računa, zabeležiti poslano vsebino ter slediti postopku organizacije za obravnavo incidentov. Preglejte veljavne možnosti brisanja in obveznosti, ne da bi obljubili, da en gumb razveljavi vse kopije ali pretekle uporabe. Nadaljujte šele po odobreni podatkovni poti. Zamenjava znamke modela brez spremembe te poti ohrani isti vzrok napake.

Končna primerjava naj vsebuje ločene presoje: izdaja je opredeljena, licenca ustreza načrtovani dejavnosti, storitev in ravnanje s podatki sta sprejemljiva, vrednotenje naloge pa je uspešno znotraj operativnega proračuna. Neizpolnjene obvezne zahteve ne moremo izravnati z lepšim besedilom ali nižjo ceno. Pri tukajšnjih primerih smo preverili dokumente in izračunali obseg pregleda; nobenega kandidata nismo zagnali ali odobrili. Tako je okvir uporaben, ne da bi si izmislili zmagovalca.

Viri

Številke virov veljajo za ta vodnik. Povezave vodijo do izvirnih dokumentov. Pred uporabo preverite besedilo, pogoje uporabe in razpoložljivost storitev.

[1] **Qwen: Qwen3-8B model card**
https://huggingface.co/Qwen/Qwen3-8B/raw/main/README.md

[2] **Qwen3-8B: Apache License, Version 2.0**
https://huggingface.co/Qwen/Qwen3-8B/raw/main/LICENSE

[3] **DeepSeek: DeepSeek-R1 model card**
https://huggingface.co/deepseek-ai/DeepSeek-R1/raw/main/README.md

[4] **DeepSeek-R1: MIT License**
https://huggingface.co/deepseek-ai/DeepSeek-R1/raw/main/LICENSE

[5] **DeepSeek Privacy Policy**
https://cdn.deepseek.com/policies/en-US/deepseek-privacy-policy.html

[6] **DeepSeek Terms of Use**
https://cdn.deepseek.com/policies/en-US/deepseek-terms-of-use.html

[8] **Fireworks: Zero Data Retention**
https://docs.fireworks.ai/guides/security_compliance/data_handling