ALTERNATOR
Misliti znanost.

Šale robotov

27. 8. 2026

Številka: 17/2026

Avtor:

- Izar Lunaček



Ilustracija: Milanka Fabjančič

© Milanka Fabjančič in ZRC SAZU, 2026. Vse pravice pridržane. Brez predhodnega pisnega dovoljenja imetnika avtorskih pravic ni dovoljeno reproduciranje, distribuiranje, javna priobčitev, predelava ali druga uporaba ilustracije v kakršnemkoli obsegu ali postopku.

Zakaj je šla umetna inteligenca čez cesto? Ker ji je človek dal tak poziv (angl. *prompt*), briga umetno inteligenco za ceste in strani.

Take robotske šale si praviloma pripovedujejo naši mehanični sluge v znanstvenofantastičnih vizijah prihodnosti in strah, ki seva iz njih, priča, da so jih napisali ljudje. Po eni strani se s takimi šalami pomirjamo, da umetna inteligenca (UI) ne more nič brez nas in da ji naše vsakdanje skrbi niso jasne, po drugi strani pa z njimi izražamo bojazen, da je ravno zato od nas boljša in da živi v neki višji, hladni sferi, kjer te skrbi sploh niso pomembne. Sam v teh spraševanjih o prihodnosti umetne inteligence vidim nekaj fantazmatskega: dišijo mi po moreči podobi starša, ki se nikoli ne zmoti in lahko zato vedno upravičeno preusmeri naše življenje ter pohodi vse naše male želje; starša, v kakršnega smo verjeli kot majhni otroci, preden smo dojeli, da ima hibe kakor mi. In upal bi si trditi, da je spraševanje o humorju umetne inteligence namenjeno prav pomiritvi te grozne vizije: če ima smisel za humor, je človeška in omejena ter ni nezmotljiv mehanizem, ki ga moramo v vsem poslušati; če pa ga nima, je preveč nečloveška in hladna, da bi ji lahko zaupali, in se lahko proti njej upravičeno borimo, da bi ohranili svoje vse preveč človeške dragocenosti.

Toda kaj se pravzaprav sprašujemo, ko se sprašujemo o umetni inteligenci in humorju? Odgovor na vprašanje, ali ima lahko umetna inteligenca dejanski smisel za humor, torej če ga lahko razume in reči dejansko ceni kot smešne, je najverjetneje ne. Občutek smešnega je po večini teorij v dobršni meri pogojen s potlačenimi impulzi, ki v njem prihajajo na dan, in algoritem za optimizacijo procesov, kar umetna inteligenca v osnovi je, teh načeloma nima – tako trdi tudi sama, če jo po tem vprašamo. Drugo vprašanje je, ali lahko umetna inteligenca sama po vnaprej danih navodilih ustvarja stvari, ki so nam ljudem smešne, in to je že bolj zanimiva tema: ali potrebujemo za duhovitost živega duha ali pa duh tako dobro ve, kaj dela, da lahko to pojasni stroju in ta duhovitost proizvaja namesto njega. A tu pride do težave: za navodila potrebujemo definicijo, definicija humorja pa je blago rečeno težavna. Ljudje skušamo humor definirati že več kot dva tisoč let, a največ, kar nam je uspelo, je, da se zdaj strinjamo, da obstajajo trije glavni sklopi definicij (https://books.google.si/books?id=GTHdPl3_chQC&lpg=PP1&hl=sl&pg=PP1#v=onepage&q&f=false) : prvi, po katerem se smejemo iz občutka vzvišenosti nad smešnim; drugi, po katerem se smejemo iz olajšanja (ob izginuli nevarnosti, a tudi ob sprostitvi neke prepovedi); in tretji, ko se smejemo zgolj nekemu formalnemu neskladju besed in misli. Preden bi torej umetni inteligenci dali navodila za izdelavo nečesa humornega, bi si torej morali izbrati vsaj definicijo, ki naj ji pri tem sledi.

Toda zdi se, da bo, tudi če si izberemo eno od definicij humorja, prišlo do zapleta. Vzemimo za primer teorijo vzvišenosti, konkretno njeno najbolj znano različico iz *Eseja o smehu* (http://www.dlib.si/?URN=URN:NBN:SI:DOC-DL313OU5) filozofa Henrija Bergsona z začetka 20. stoletja. Bergsonova ideja je, da se smejemo članu družbe, ki je za hip nehal biti gibek, sam sebe opazujoč in živ ter je postal mehaničen in ponovljiv, s smehom pa ga za to blago grajamo, ker prekršek ni dovolj velik za resno družbeno sankcijo. Kar je pri tem zanimivo, je, da so teme, ki jih pri smehu odpira Bergson, povezane ravno z razmerjem med živim in mehaničnim, strojnim in človeškim, skratka teme, ki nas zanimajo tudi pri umetni inteligenci. In ravno tu pride do težave: če smo smešni, ko kot živa, organska, načeloma gibka bitja delujemo rahlo mehanično, togo in strojno, kako naj ta obrat poustvari neki mehanizem, načeloma namenjen maksimalno gladkemu teku? Idealno delujoča UI bi bila ravno tako, samo sebe popolnoma opazujoče in kar naprej popravljajoče se življenje, ki ga Bergson postavlja kot ideal, ljudje pa ga ne dosegamo vedno. Smešna bi bila, ko bi se zmotila – in s tem postala, po Bergsonu, »mehanska«, ne dovolj življenjska.

Popkultura je polna motivov robotov, ki se pokvarijo in šele takrat postanejo človeški, in to na komičen način (*Wall-E*, *Short Circuit*), a to je pač človeška fantazma o UI, pomirjujoča ob nasprotni predstavi umetne inteligence kot brezčutno učinkovite (Wall-E proti robotu krmarju, zasnovanem na HAL 9000 iz *Odiseje 2001*). Toda ali bi lahko bila UI komična zanalašč – kot komik, ki načrtuje svoje spodrsljaje, da so videti nehoteni? Zdi se, da je ravno tu težava UI: komika živi prav v zglobu med strojnim in živim ter med hotenim in nehotenim, točno tam, kjer je doma tudi UI (daje nam, kar smo hoteli, a ne povsem; je neživa, a hlini živost), le da pri UI to skoraj vedno izpade rahlo grozljivo namesto smešno.

Teorija vzvišenosti je najbolj prikladna za telesno komiko, medtem ko so najbolj priljubljena sodobna umetna inteligenca veliki jezikovni modeli, zato bomo imeli mogoče več sreče, če jih za produkcijo smešnega opremimo s teorijo neskladja. Po njej se smejimo, ko se neki koncepti ali besede sestavijo na neprimeren način, se pravi ne zaradi notranje logične povezanosti, ampak na podlagi nekih bolj površinskih stikov, kot je zven besed ali nebistven stik konceptov. In res na tej podlagi deluje doslej najbolj uspešna UI na področju humorja, tj. generator šal Witscript (https://undark.org/2025/07/21/ai-humor) , ki ga je lani razvil pisec dovtipov za ameriške nočne pogovorne oddaje in avtor priročnika za ustvarjanje takega programa. Witscript deluje tako, da mu poveste naslov iz časopisa, on pa zanj ustvari šalo tako, da v naslovu identificira dva ključna koncepta, za vsakega razvije nabor asociacij, nato pa oboje spoji v nov stavek, ki ob pravi kombinaciji logičnega in absurdnega deluje komično. Šalo mora nato v ponujenem naboru še vedno izbrati človek, a vsaj en, razmeroma mehaničen del procesa zanj opravi stroj. Zanimivo je, da mora šalo tudi na odru povedati človek, saj občinstvo šale, za katere ve, da jih je napisala UI, ceni manj. Tu gre deloma verjetno za človeško fetišizacijo truda (pri balerini cenimo, če neko pirueto izvede skoraj tako natančno kot stroj, če pa je res stroj, nas to ne navduši), deloma pa še za nekaj drugega, povezanega s tem, kar smo prej omenjali pri telesni komiki – da je tudi zdrs ali napaka vsaj v humorju cenjena bolj, če jo izvede človek, pri katerem v nasprotju s strojem ne vemo zagotovo, ali je bil zdrs hoten ali nehoten.

To pa je povezano s tretjim sklopom teorij komičnega, tistih, ki trdijo, da se smejemo ob sprostitvi, zlasti sprostitvi neke prepovedi. Tu je verjetno najbolj vplivna teorija Sigmunda Freuda (https://plus.cobiss.net/cobiss/si/sl/data/cobib/125178624) , ki trdi, da vic (torej ena od zvrsti besedne komike) preko na videz nedolžnega igranja z besedami in koncepti obelodanja prepovedane, potlačene vsebine, ki so v resnici glavni vir našega užitka v smehu. Freud priznava, da obstajajo tudi povsem nedolžni vici (na internetu kar mrgoli slabih besednih iger (angl. *bad puns*) in očetovskih šal (angl. *dad jokes*) in v teh je UI, vključno z Witscriptom, res še najboljša), a ti nas še zdaleč ne zadovoljijo tako kot vici s prikrito obsceno vsebino, kjer se – po Freudu – sicer lahko pretvarjamo, da uživamo v pametni, duhoviti embalaži, a v resnici nas veseli predvsem to, kar skoznjo prepovedanega prihaja na dan. Seveda bi bilo UI še težje razložiti, kaj je v danem trenutku ravno prav prepovedano, da bo povzročilo maksimalni užitek, a tudi če bi se ta tega lahko skozi opazovanje situacij naučila sama, se zdi, da bi tudi ob popolni obsceni šali, ki bi jo ustvarila UI, ljudi nekaj zmotilo. Namreč: zdi se, da ljudje cenimo, če uspe prepoved z uporabo lastne duhovitosti (angl. *wit*) zaobiti nekomu, ki si z nami deli te blokade in ki s tem kot ranljiv posameznik na odru tvega naš linč. Užitek v šali je toliko močnejši, kolikor bliže roba gre, kolikor bolj tvegana je, kolikor bolj napne razmerje med nesprejemljivo, obsceno vsebino in duhovito, pametno formo, s katero jo komik pretihotapi mimo cenzorja.

Tudi če torej jasna definicija humorja obstaja, pa ne moremo reči, naj umetna inteligenca deluje po njenih navodilih, ker je bistvo transakcije v tem, da ne sme biti odkrito jasno, da nas nekdo zabava po naših navodilih. Ker v humorju kršimo družbene norme, moramo biti v smeh vsaj na videz nekako porinjeni, prisiljeni, in celo producent humorja se mora sam pred sabo delati, da ne ve povsem, kaj počne, in da tudi njega samega presenečajo obsceni smisli njegovih igric. Če je humor res siva cona, v kateri lahko razkrivamo prepovedane vsebine, ki se v resen diskurz (še) ne smejo prebiti, potem tega ravno nočemo definirati, ker je bistveno prav to, da je to polje nedefinirano, nezasedeno z jasnim, racionalnim smislom in namenom.

Umetna inteligenca bo torej težko postala polnokrvno humorna. Lahko se nauči nedolžnih besednih iger, morda bi se lahko naučila tudi dovolj dobro prepoznati zapletene psihosocialne sisteme v vsakem hipu dovoljenih, vzpodbujanih in prepovedanih vsebin ter kako se z njimi igrati, da bi lahko ravno prav prepovedane obelodanila skozi ravno prav vzpodbujane, toda za popoln učinek bi vsekakor potrebovali občutek, da je ta, ki nam govori šale, tudi sam ranljiv, tako za prepovedi kot za naš gnev ob njihovi prekoračitvi. Očetovske šale se tako imenujejo z razlogom: prihajajo s točke, ki jo z očmi zavijajoči najstnik dojema obenem kot točko oblasti in irelevantnosti, pomanjkanja stika s svetom podrejenih. UI je v odnosu do človeštva stalno v nevarnosti, da bo razumljena kot tak neprimerno šaleči se očka.

In tu, v zaključku, prihajamo do zares ključnega konflikta med projektom UI in humorjem. UI se, kot je zasnovana sedaj, to je kot orodje za optimizirano avtomatizacijo vseh človeških dejavnosti, tako prijetnih kot neprijetnih, zdi dokaj nezdrava za

človeško psiho. Zgodovinsko gledano smo stroje razvili, da bi nas odrešili zoprne tlake in da bi si sprostili čas za bolj kreativne, človeškega duha vredne dejavnosti. Že tu smo se vsaj malce ušteli, ker opažamo, da ljudje, ki so povsem brez ročnega dela, tudi rahlo slabijo. Toda če na stroje preložimo še prijetne, kreativne reči, se res znajdemo v težavah. Zakaj?

Prvič zato, ker je kreativne reči prijetno početi, zato jih je nesmiselno prelagati na stroje. Tu se je že pojavil ugovor neumetnikov, češ da se je z UI kreativnost demokratizirala, kot se je s fotoaparatom demokratizirala odslikava realnosti, ki je bila prej domena slikarjev. Toda do kreativnosti ne obstaja te vrste bližnjica: da bi nas na dolgi rok res zadovoljevala, mora vključevati tudi učenje obrti. Obrt je lahko tudi obvladanje UI in ljudje, ki jo res obvladajo, jo uporabljajo za nove vrste kreativnosti – tako, kot bi moralo biti. Problematično pa je, kadar skušajo racionalni menedžerji prek UI prekiniti sodelovanje z umetniki, ker je usklajevanje šefove vizije s kreativčevo pregovorno naporno. Na več mestih v razvoju UI je mogoče kot motivacijo čutiti določen gnus do sočloveka, med drugim tudi v na videz velikodušnem opremljanju robotskih družabnikov s smislom za humor, s čimer naj bi se pomagalo osamljenim ljudem. Iz te predstave veje gnus tako do osamljenih ljudi, s katerimi se končno ne bo treba več ukvarjati, kot tudi gnus osamljenih ljudi do drugih, ki jih bo končno nadomestil robotski služabnik.

Drugič: UI kot stroj, ki zna tudi človeške reči, poglablja naš izvirni odpor do soljudi (ki so ga radi omenjali tako eksistencialisti kot naš domači Žižek), ki se je z individualizmom zadnjega stoletja že tako in tako razrasel in verjetno ne potrebuje dodatne vzpodbude.

In tretjič, če stroju prepustiš ne le neprijetne opravke, za katere se mu lahko poda jasna navodila (prav zato so to neprijetni opravki), ampak tudi tiste, za katere se ne da podati jasnih navodil, se bo ta moral pravilnega izvajanja učiti skozi reakcije. Pri tem nas bo pahnil v isto past kot algoritmi za priporočila, ki že delujejo po tem principu: iz vzorcev bo razbiral naše želje in nas potiskal še globlje v mehurčke lastnih fetišev in preferenc. Sodobna UI ni navaden stroj, ki zgolj avtomatizira in optimizira neko našo mehanično dejavnost (kar je bil tudi fotoaparat za slikarstvo); priliznjen parazit je, ki prisluškuje našim željam in jih krepi, streže jim tako, da vase srka tisočletja človeške kreativnosti in jih vrača prežvečene v »dovolj dobre« rešitve, da imamo občutek, da umetnikov – in soljudi nasploh – za realizacijo svoje vizije ne potrebujemo. UI, kot je zastavljena zdaj, je problematična, ker je le podaljšek algoritemsko optimizirane uporabniške izkušnje interneta in ker nas po isti poti kot ta dela neodvisne od drugih ter odvisne od odmevnih komor lastnih preferenc.

In morda se rešitev skriva prav v humorju. Dober humor namreč vedno ve, da je želja naluknjana in da je ni brez Drugega. Da se prav na bolečih presekih različnih resničnosti skriva užitek presežka. Da takrat, ko hočemo ravno prav, vedno dobimo preveč. In da gnusnosti sočloveka ne moremo zares uiti: če se ji še tako izmikamo, na koncu ostanemo sami s svojo lastno. Zato nam bo humor v tem novem poglavju človeške zgodovine prišel prav, če se bodo optimizirajoči algoritmi res zalezli v vse pore našega življenja in če bomo postali vase zaprti sledilci lastnih preferenc, ki jih bo pred vsemi nevarnostmi drugačnosti obvaroval ta dobrohotni in dušeči starš. V filmu *Wall-E* je ljudi na koncu pred tako usodo s kratkim stikom rešil naslovni robot, ki je premagal dušečo UI in ljudi zbezal iz naslanjačev – v resničnem scenariju bo rešitev morala priti iz nas, če pa bo prišla od zunaj, najverjetneje ne bo tako prijazna kot omenjeni robotek. Bodimo torej sami sebi Wall-E in se z lastno mehanično komičnostjo odtrgajmo od pretirane odvisnosti od strojev ter v bolečem stiku z gnusnimi drugimi znova ukrešimo iskro svoje duhovitosti, ki jo potrebujemo za nove iznajdbe, nove vizije in nove zmage.

* Besedilo je prvotno izšlo v tiskani tematski številki // Preživetje človeštva, ki jo lahko naročite tule (https://knjigarna.zrc-sazu.si/sl/publikacije/alternator-2026) , naprodaj pa je tudi v knjigarni Azil (https://maps.app.goo.gl/kmCnd1G1fuWnLZV76) v Ljubljani in knjigarni Maks (https://maps.app.goo.gl/wNHriAa3xtCgpSFq5) v Novi Gorici.