Neznanje u veštačkoj inteligenciji

Damjan Krstajić

Ako vam neko pokaže sliku životinje koju nikad do sada niste videli, ni čuli za nju, i pita vas da je imenujete, vi biste odgovorili jednostavno „*Ne znam*“. Međutim, ako to isto uradite u (onom što neki nazivaju) sistemu veštačke inteligencije koji prepoznaje životinje na osnovu slika, dobili biste kao odgovor neku ranije poznatu životinju sa najvećom verovatnoćom da liči na onu sa slike. Nema „*Ne znam*“. Zašto?

Jednostavno, u matematičkoj verovatnoći i statistici ne postoji odgovor „*Ne znam*“. Postoje neslaganja među vrhunskim svetskim matematičarima u vezi sa tačnom definicijom verovatnoće nekog događaja A, ali se svi slažu da je to mera sa vrednostima između 0 i 1 (kolokvijalno između 0% i 100%), gde 0 označava da je u pitanju nemoguć događaj, dok 1 (100%) da će se sigurno desiti. Primera radi, ako bacamo novčić, verovatnoća da padne pismo je 50% i ista je kao da padne glava. Pojednostavljeno, verovatnoću događaja A možemo da posmatramo kao odnos argumenata za i protiv da će se A desiti. Međutim, šta da radimo u slučaju događaja B, nešto potpuno novo kao, na primer, slika nepoznate životinje, za koji nemamo nijedan argument ni za ni protiv? Koja je verovatnoća događaja B? Moj odgovor je „*Ne znam*“.

Većina onog što danas čujemo u vestima da se naziva veštačka inteligencija donosi odluke koje se baziraju na verovatnoćama. Po mom mišljenju, osnovna mana tih sistema je što većina njih nema ugrađen odgovor „*Ne znam*“. Zašto? Nemam odgovor na ovo pitanje, ali pretpostavljam da kad vam neko prodaje sistem koji naziva da je veštačka inteligencija, obično se fokusira na to šta taj sistem sve može, a ne na to šta sistem ne zna. Iz mog iskustva, generalno odgovor „*Ne znam*“ nije uopšte popularan u akademskoj zajednici, a nije ni interesantan medijima, ali jeste veoma važan u realnom životu.

Nedavno je britanski *Royal Society of Chemistry* objavio knjigu „*Artificial Intelligence in Drug Discovery*“ u kojoj je autor ovog članka bio pozvan da napiše poglavlje o dobrobiti definisanja neznanja u veštačkoj inteligenciji. Ovde ću izložiti suštinu mojih argumenata, a tehnički detalji mogu da se nađu u knjizi.

Prvo, odgovor „*Ne znam*“ je jako bitan zbog poverenja. Neka imamo dva sistema veštačke inteligencije X i Y, gde X zna više od Y, ali X ne ume da odgovori „*Ne znam*“, dok je Y sposoban za to. Kome biste više verovali?

Drugo, ako se budemo oslanjali na sistem veštačke inteligencije u važnim odlukama, onda naše akcije mogu biti drugačije ako postoji odgovor „*Ne znam*“ od strane veštačke inteligencije. Trenutno u medicinskoj dijagnostici postoje razni predikcioni sistemi koji na osnovu krvne analize ili rendgen snimka informišu korisnika da li neki pacijent ima tumor ili oboljenje i sa kojom verovatnoćom. Recimo da je sistem dao 45% šanse da pacijent ima tumor. Kako da razumemo ovaj rezultat? Psiholog Danijel Kaneman (Daniel Kahneman), dobitnik Nobelove nagrade za ekonomiju, pokazao je zajedno sa Amosom Tverskijem (Amos Tversky) kako ne samo obični ljudi, već i profesionalni statističari, mogu da pogreše u razumevanju verovatnoća u svakodnevnom životu. Ukratko, verovatnoću od 45% neki doktori bi protumačili na jedan način, a drugi drugačije, i samim tim njihove akcije bi bile različite. Međutim, pretpostavljam da bi svi lekari isto razumeli odgovor „*Ne znam*“ i po mogućstvu zatražili dodatna pretraživanja. Samo da napomenem da u svom čitanju Kanemana i Tverskog nisam naišao na njihova istraživanja da li ljudi reaguju isto na odgovor „*Ne znam*“.

Treće, retko ko vidi da će u budućnosti veštačka inteligencija biti fiksan i nepromenljiv sistem. Drugim rečima, posle kreiranja, sistemu veštačke inteligencije će biti potrebno podučavanje u hodu. Odgovor „*Ne znam*“ od strane veštačke inteligencije je tada bitan, jer će nam omogućiti da znamo šta sistemu nedostaje i moći ćemo da ga unapredimo.

Po meni, odgovornost kreatora sistema veštačke inteligencije je da informišu korisnike o ograničenjima sistema sa odgovorom „*Ne znam*“. Oni su, siguran sam, i sami svesni da njihove kreacije imaju limite, ali mi nije poznato da ih neko tera da ih definišu. Problem je donekle i ljudske prirode, jer nama odgovor „*Ne znam*“ dolazi nekako spontano, kao rezultat introspekcije. Nikad se ne pitamo kako znamo da nešto ne znamo. Jednostavno je tako. Međutim, problem kako kreirati veštačku inteligenciju da zna kad ne zna je mnogo kompleksniji i teži, jer veštačka inteligencija nema lukzus introspekcije.

Mišljenja sam da progres u istraživanjima i razvoju veštačke inteligencije zavisi od kreiranja sistema koji su sposobni (figurativno rečeno) da znaju kad ne znaju. Verujem da to nije samo moguće, već i odgovornost kreatora. Bez upoznavanja sa ograničenjima veštačke inteligencije, koja su

ključna za poverenje, dolazimo u opasnost da slepo verujemo u veštačku inteligenciju i tako joj na duge staze oduzimamo šansu za veći uspeh.

Reference koje podržavaju činjenice spomenute u članku

1. Knjiga *Artificial Intelligence in Drug Discovery*

https://www.amazon.com/Artificial-Intelligence-Drug-Discovery-ISSN-ebook/dp/B08MYSRPSF/

https://pubs.rsc.org/en/content/ebook/978-1-78801-547-9

2. Moje poglavlje *Non-applicability Domain. The Benefits of Defining „I Don‘t Know“ in Artificial Intelligence*

https://pubs.rsc.org/en/content/chapter/bk9781788015479-00102/978-1-78801-547-9

3. Danijel Kaneman (Daniel Kahneman)

https://en.wikipedia.org/wiki/Daniel_Kahneman

4. Amos Tverski (Amos Tversky)

https://en.wikipedia.org/wiki/Amos_Tversky

5. Kahneman i Tverski su opisali kognitivnu pristrasnost

https://en.wikipedia.org/wiki/Cognitive_bias