Elena LAZKANO ORTEGA
Donostiako EHU-ko Informatika Fakultateko irakasle eta ikerlaria

Zein da zein, Pepper?

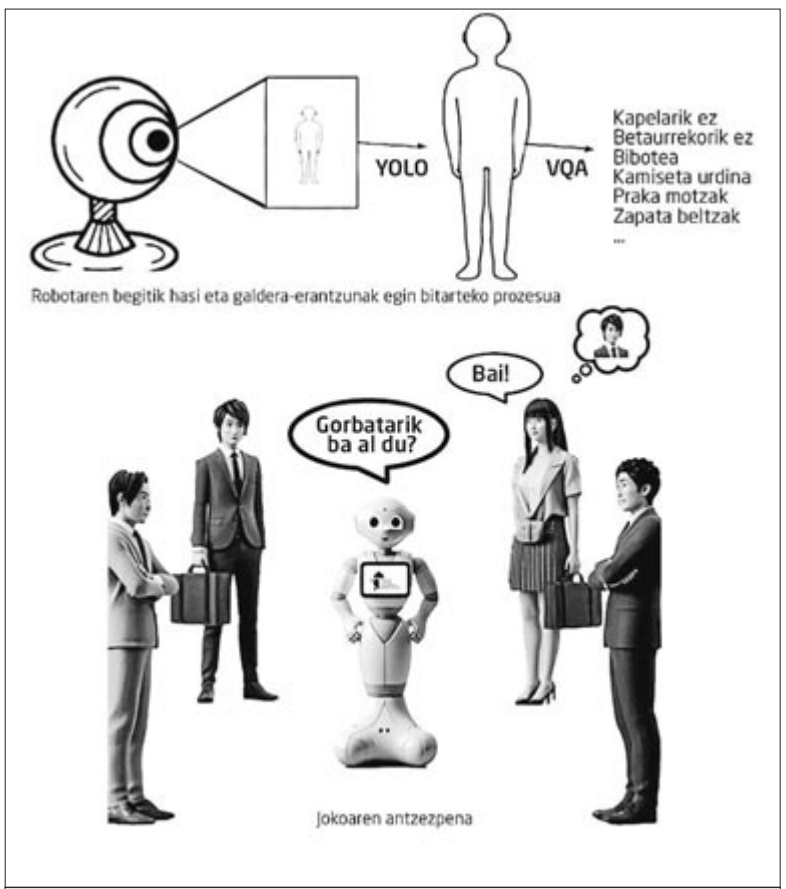
Nabarmen, robot sozialek elkarrekintza natural eta pertsonalizaturako gaitasuna erakutsi behar dute, eta horretarako, ezinbestekoa da aurrean duten hori nor den ezagutzea. Epe luzerako elkarrekintzen kasuan, adibidez, robota etxean izango badugu laguntzaile gisa, ezaugarri biometriko gogorrak erabili behar dira ezinbestean: begien kolorea, aurpegiaren forma, aurpegiko beste zenbait marka... hauek datu sentsibleak dira, eta pribatutasun kezkak sortzen dituzte. Baina epe motzeko elkarrekintzetarako, horren intrusiboak ez diren atributuez balia gaitezke, ezaugarri biometriko bigunez, alegia: nolako jertsea dugun soinean, praka motzak edo luzeak daramatzagun, zapata gorriak edo beltzak ditugun, betaurrekorik bai edo ez, etab. luzea. Nola harrapatu hauek? Adimen artifiziala erabiliz, noski!

Ikusizko hizkuntza-ereduek (Visual language Model, VLM) irudi edo/eta testuaren gainean arrazoitzeko gaitasuna erakusten dute, ez dira berriak, hamarkada bete dute jada. Lehenengo ereduek sare konboluzionalak erabiltzen bazituzten irudietatik ezaugarriak eratortzeko, gaur egun hizkuntza-eredu handi (Large Language Model, LLM) modernoetan daude oinarrituta, eta arrakasta handia izaten ari dira. Irudien edukia testu moduan azal dezakete, irudien inguruko galderei erantzunez (Visual Question Answering, VQA) edota azalpen batekin bat datorren irudiak bilatuz, besteak beste.

Ikusizko galdera-erantzun (VQA) ereduak, berez, VLM mota bat dira. Konputagailu bidezko ikusmena eta lengoaia naturalaren prozesamendua konbinatzen dituzte irudi batek gordetzen duen informazioa ulertu eta interpretatzeko. Haien ezaugarri behinena, galdera motzei erantzuteko prestatuta egotea da. VQA ereduek errendimendu lehiakorra erakuts dezakete berezko entrenamendurik gabe, "zero-shot" moduan. Ez da ezaugarri makala hau, datu-base erraldoiak eraiki, etika batzordeak aztoratu eta, areago, txartel grafikoz jositako ordenadoreak erretzen baitabiltza gaur egun AA orokorraren lehian dabiltzan agenteak, ereduak entrenatu eta egokitzeko.

Ikusizko galdera-erantzun eredu batez baliatu gara gure Pepper robota aski ezaguna den "Zein da zein?" jokoan aritu dadin. Pepper-ek inguruan dituen pertsonak desberdin ditzake BLIP-2 ikusizko galdera-erantzun sistema aurre-entrenatuaren bidez. Fitxarik ez dugu erabiltzen, joko bizia baizik. Korro bat eginez robota inguratzen duten pertsonek hartzen dute parte, eta robota eta jokalari bat izango dira asmatzailea bata eta jokoaren gidaria bestea, txandaka. Gidaria gizakia zein Pepper izanda, robotak pertsonak kokatu behar ditu bere inguruan, biratu ahala begi-kolpeak eman eta kameratik jasotzen duen irudian pertsonarik dagoen erabaki. Azken funtzio hori YOLO1-k beteko du. Gero, gidariaren kasuan, korroko pertsona bat aukeratu beharko du eta asmatzailearen galderak erantzun beharko ditu; edo beste aldera, galderak egin beharko ditu eta inguruko pertsonak baztertzen joan, jasotako erantzunen arabera.

Etorkizun hurbilean, agian etxean izango duzun robotak esango dizu ea zure janzkera bat datorren eguraldiarekin. •



Robotaren begitik hasi eta galdera-erantzunak egin bitarteko prozesua.