

# Jakin-minaren indarra

Zerk bultzatzen gaitu gure mundua esploratzera? Zergatik nahi du ukitu haurtxo batek inoiz ikusi ez duen baloi hori? Non aurkitu daiteke gure historiako bidaiari handien motibazioa? Zer zerabilten buruan Ilargira joan ziren astronautek? Zergatik tematzen dira zientzialariak ezezaguna dena azaltzen? Geroz eta ziurrago dakigu jakin-mina dela gure garuneko motorra martxan jartzen duen erregaia.

Azken 15 urteotan giza garunaren funtzionamenduaren inguruan egin diren hainbat esperimentu eta saiakerek antzeko ondorio bat ateratzera garamatzate: etorkizuna aurreratzeko makinak dira garunak. Korrante hori ongien azaltzen duenatariko bat Andy Clark filosofoa da [1]. Haren arabera, giza pertzepzioa kontrolpean gertatzen den haluzinazioa besterik ez da. Hots, ez dugu errealtatea den bezala hautematen; gure garunak irudikatu egiten du, haluzinatu.

Har dezagun pilotari bat. Sake bat errestatu behar duenean, adibidez, pilota 100 km/h-ko abiaduran joan ohi da. Beraz, pilotariak milisegundo batzuk ditu erabakitzeko nola jo behar duen pilota. Lan txarrak dira horiek, begietatik garunera seinaleak iristeko denbora gehiago behar baitu. Baina pilotariaren garunak, esperientzian oinarrituta, aurrean dezake pilota non egongo den une oro. Garunaren estimazio horretan oinarrituta mugitzen da bere gorputza, eta behar bezala jo dezake pilota.

Andy Clarkek defendatzen duenaren arabera, gure garunak jaiotzetik trebatzen dira etorkizuna aurreratzeko. Hala, errealtatearen ereduak sortzen ditugu, munduaren dinamika ulertzen duten ereduak. Eta eredu horiek erabiltzen ditugu erabakiak hartzeko.

Idea horiek baliozkoak direnez ikusteko, besteak beste, adimen artifiziala erabil daiteke. Hori bera pentsatu zuten David Ha eta Jürgen Schmidhuber ikertzaileek [2].

## Mundu-ereduak eta errefortzu bidezko ikasketa

Errefortzu bidezko ikasketa animaliekin erabiltzen den ikasketa-modu bat da. Oso oinarri sinpleak ditu. Adibidez, zakur bati “eseri” esatean eser dadin nahi badugu, irakatsi egin beharko diogu. Horrela, “eseri” esan eta ez badu egiten guk nahi duguna, *zigortu* egingo dugu. Baina esertzen bada, *sari* bat emango diogu. Sariak eta zigorrak jasoz, zakurrak ikasi egiten du zer egin behar duen “eseri” agindua entzuten duenean.

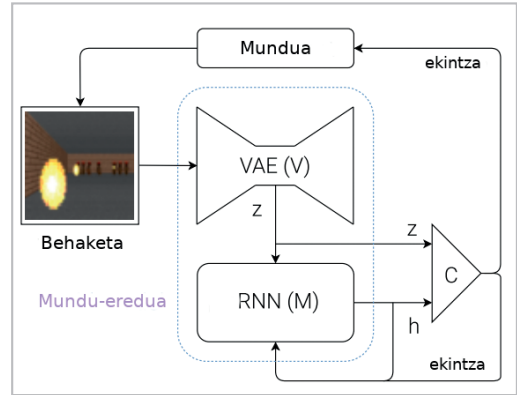
Adimen artifizialaren munduan ere errefortzu bidezko ikasketa erabiltzen da, batzuetan arrakasta handiz, gainera [3]. Oso modu laburrean, 1. irudian ikus daiteke errefortzu bidezko sistema batek nola funtzionatzen duen.

Baina errefortzu bidezko soluzio klasikoek bi arazo nagusi izaten dituzte:

1. Ataza jakin baterako bakarrik balio dute. Ataza bat egiten ikasitakoa ezin da beste ataza batean aplikatu, nahiz eta bi atazen artean antzekotasun handiak egon.
2. Errefortzua jaso behar izaten dute ia une oro. Mundu errealeko ataza askotan, ezin da errefortzua jaso ataza amaitu arte. Esaterako, makina bat trebatu nahi badugu karta batzuekin gaztelu bat eraikitzeko, ezin diogu errefortzurik eman gaztelua bukatu artean.

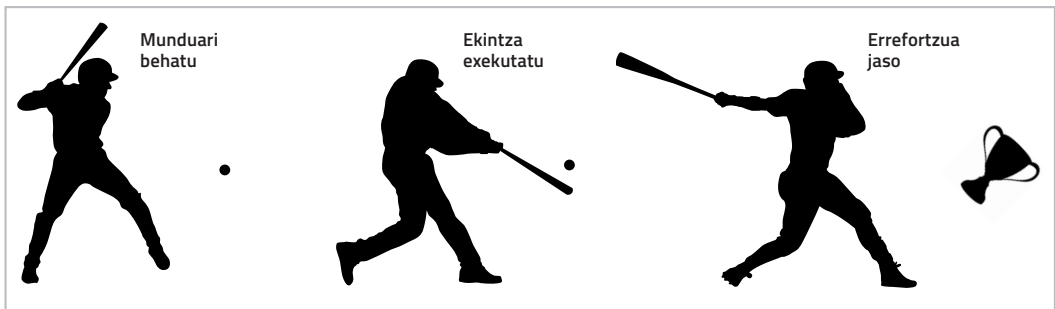
Arazo horiek gainditzeko, mundu-ereduak ikas-  
teko gai den sistema bat proposatzen dute Hak  
eta Schmidhuberrek (2. irudia). Azal dezagun labur  
haien ideia zein den: makina batek behaketa bat ja-  
sotzen duen bakoitzean (kamera bidezko irudi bat,  
adibidez), jaso duen informazio hori errepresentatu  
behar du lehendabizi. Horretarako, sare neuronal  
berezi batzuk erabiltzen dituzte: ingelesez, Varia-  
tional Auto-Encoders (VAE) deitzen zaie (2. irudiko  
V modulua). Pentsa dezagun kamera bidezko irudi  
batean agertzen diren ezaugarri garrantzitsuenen  
laburpen bat sortzen duela VAE horrek (2. irudiko  
Z, non Z zenbakien bektore bat den). Une horretan  
jasotako behaketa hori M moduluen sarreretako  
bat bilakatzen da. M modulua beste sare neuronal  
bat da (kasu honetan, sare errekurtsibo bat: RNN).  
Une honetako behaketa (Z), exekutututako azken  
ekintza eta memoria erabiliz (h), makinak ikusiko  
duen hurrengo behaketa zein izango den aurre-  
satea du helburu M modulua. Pilotariaren kasuan,  
M modulua esango luke pilota non egongo den  
etorkizunean, pilotariaren azken mugimenduak eta  
munduaren dinamika kontuan hartuz.

Amaitzeko, une honetako behaketa (Z) eta etor-  
kizunerako espero dugun behaketa (h) abiapuntu  
harturik, kontrolagailu simple bat erabiltzen da (C)

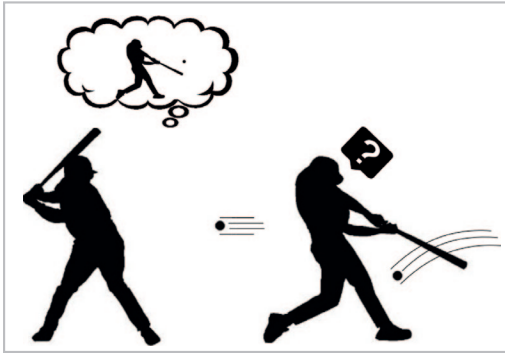


2. irudia. Mundu-ereduen proposamena. Ikusmen bidezko  
behaketa jaso (V), eta haren errepresentazioa sortzen da (Z).  
Erreprentazio hori memoria-modulura eramaten da (M);  
horren helburua da hurrengo pausoa zer behaketa  
ikusiko den aurrestea, oraingo behaketa eta azken ekintza  
kontuan hartuta. Oraingo behaketaren erreprentazioa eta  
aurreteko behaketa erabiliz, kontrolagailuak (C)  
erabakitzen du zer ekintza exekutatu behar den.

makinaren hurrengo mugimendua zein izango den  
erabakitzeko. Kontrolagailua, kasu honetan, sare  
neuronal simple bat da, eta errefortzu bidez ikasten  
du. Baina mundu-eredua (V eta M moduluak) ez da  
errefortzu bidez entrenatzen, eta, beraz, ez du lotu-  
tarik ataza jakin batekiko. Teorian, egin nahi ditugun  
atazekiko modu independentean eraikitzen ditu-  
gu gizakiok mundu-ereduak, eta baliagarriak dira  
ataza guztietarako. Pilota baten dinamika ikasten  
dugunean, hots, pilotaren mugimenduaren eredia  
sortzen duenengan gure garunak, gai gara eredu  
hori kirol askotan erabiltzeko: pilotan, tenisean, fut-



1. irudia. Errefortzu bidezko ikasketa-sistema baten elementu nagusiak eta prozesua. Munduari behatzen zaio (pilotaren  
ibilbidea), eta behaketa hori ikasketa-sistemara bideratzen da (bateatzailea). Behaketetan oinarrituta, sistemak erabakiak  
hartzen ditu, eta ekintzak exekutatu (kasu honetan, pilota jotzeko aldaera jakin bat). Ekintza horien ondorioen eta ikasi nahi  
den atazaren arabera, errefortzu-funtzioak errefortzua emango dio ikasketa-sistemari (pilota ongi jotzeagatik bateatzaileak  
saria jasoko du). Hala ikasiko du sistemak ataza gauzatzen.



3. irudia. Bateatzaileak pilotari behatzean, pilotak etorkizunean zer ibilbide izango duen aurreساتen du, eta mugimenduak planifikatzen. Exekutatzean, ordea, konturatzen da pilotak ez duela egin berak espero zuen ibilbidea. Demagun efektuz botatako pilota bat ikusten duen lehen aldia dela. Aurreساتandako eta benetakoa ibilbideak oso ezberdinak direnez, bateatzaileak ezusteko handia hartuko du.

bolean eta saskibaloian, esaterako. Kirol bakoitzean exekutatu beharreko ekintzak ezberdinak dira, arau eta helburu ezberdinak baitituzte, baina baloia alda pilotaren dinamika bera da. Beraz, kirol bakoitzerako kontrolagailuak ikas daitezke erreformatu bidez, baina mundu-ereduak partekatuz.

Nola trebatzen dira, bada, mundu-eredu horiek? Etorkizuneko behaketak aurreساتen, besterik gabe. Sare neuronalen kasuan, une bateko behaketa bat emango genieke V eta M moduluei, eta hurrengo uneko behaketa iragarri beharko lukete. Hau da, haien helburua da aurreساتen duten behaketa eta ondoren ikusten dutenaren arteko aldea minimizatzea. Eta horrela ikasten dute geroz eta iragarpen hobeak egiten. Kontu egin ikasketa horrek ez duela zerikusirik ataza jakin batekin.

### Jakin-mina formulatzen

Demagun, beraz, badakigula jada nola sortu mundu-eredu bat edozein atazatarako. Laguntzen al digu jakin-mina ulertzen etorkizuna aurreساتeko gai den mundu-eredu horrek? Andy Clarken arabera, gure garunak etorkizuna aurreساتen du, baina ezusteko

handia hartzen du aurreساتandako etorkizun horrek ez badu zerikusirik benetan behatzen duenarekin. Ezuste horiek adierazten dute gure mundu-eredua ez dela behar bezain ona. Beraz, gure garunak arreta berezia eskaintzen die une horiei. Behaketa arraro horiek aztertzen ditu, hainbat ekintza exekutatzeko, eta bere ezustearen zergatia bilatu nahi izaten du (3. irudia). Ez al da, bada, hori jakin-mina? Bingo!

Era horretara esanda, badirudi ez dela hain zaila jakin-mina formulatzea: konpara dezagun mundu-ereduak sortzen duen iragarpena benetan ikusitakoarekin; antzekoak badira, egoera hori ezaguna da eta, beraz, ez dugu gehiago aztertu behar; baina oso ezberdinak badira, piz dezagun gure jakin-mina. Erreformatu bidezko ikasketaren ideiak jarraituz, jaso dezala makinak saria bere ekintzen ondorioz ongi aurreسات ezin dituen egoerak topatzen dituen bakoitzean. Eta horrela, makinak bere ingurunearekin jolastuz ikasiko du, ume txikiek egiten duten antzera. Era honetara formulatzen den erreformatua *berezko erreformatu* ala *motibazio* gisa ezagutzen da.

### Jakin-minak gidatutako ikasketa

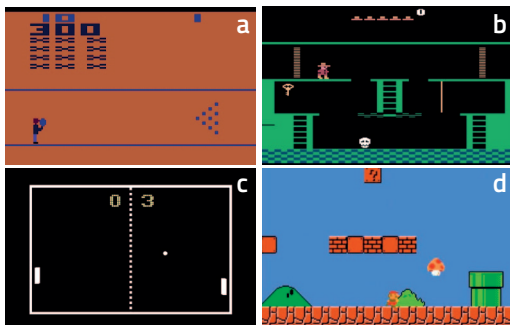
OpenAI ikerketa-zentroak jakin-minaren indarra aztertu nahi izan zuen [4]. Horretarako, bideo-joko klasiko pilo bat hartu zituzten, eta haiekin entrenatu zituzten beren makinak. Etorkizuna aurreساتeko gai diren mundu-ereduak zituzten oinarritzat makina horiek. Eta lehen esan dugun bezala formulatzen zuten jakin-mina. Hots, iragarritakoa eta benetan behatutakoaren arteko aldeak kontuan hartuta.

Aipatutako joko horiek helburu ezberdinak dituzte, eta erreformatu-funtzioan islatzen dira helburuak. Baina erreformatu hori alde batera utzi zuten OpenAI-ko ikertzaileek, jakin-mina bakarrik erabiliz. Beraz, makinek sariak jasotzen zituzten beren ekintzek egoera berrietara eramaten zituztenean: aurreساتeko zailak ziren egoeretara, hain zuzen.

Esan daiteke motibazio hutsez trebatutako makinak zirela.

Emaitza txundigarriak lortu zituzten. Adibidez, Bowling jokoak (4.a irudia) jakin-min hutsez gaindizten ikasi zuen makinak. Ikertzaileen arabera, pantailan puntuazioa agertzea da emaitza hori azaltzeko gakoa. Birla asko botatzen baditu makinak, puntuazioa asko aldatzen da. Beraz, hasiera batean, egoera horiek dira auresaten zailenak, hots, behaketa aldaketa handienak eragiten dituztenak. Pong-en kasuan, bestalde (4.c irudia), pilota denbora luzez jokoan mantentzen ikasten du makinak, baina ez irabazten. Hala ere, argi dago bikain ikasten duela munduaren dinamika. Super Mario Bros-en kasuan ere (4.d irudia), jokoak gaindizten ikasten du jakin-min hutsez.

Harrigarriena, hala ere, Montezuma's Revenge jokoan gertatu zen (4.b irudia). Joko hori oso konplexua da makinentzat gaur egun, esplorazio-gaitasun handiak eskatzen baititu eta errefortzuak oso gutxitan jasotzen baitira. Gaur egun hain modan dauden *escape room*-en antzekoa da joko hori: puzzleak osatzen joan behar da, gela ezberdinak osatu eta jokoak gainditzeko. Errefortzuak gela bakoitzetik irtetean iristen dira (irtenez gero, saria; bestela, zigorra). Orain arteko makinek ezin izan dute gizakion puntuazioetara hurbildu, ez baitzuten errefortzurik



4. irudia. OpenAI-ren esperimentuetan erabilitako lau bideo-joko: a) Bowling, b) Montezuma's revenge, c) Pong eta d) Super Mario Bros.

jasotzen gela bakoitzeko objektu ezberdinak esploratu eta puzzleak askatzen saiatzeagatik. Jakin-min hutsez trebatutako makinak, berriz, gizakion antzeko emaitzak lortu ahal izan zituen! Azken batean, objektuak eta gela bakoitzeko txoko arraroak esploratzera bultzatzen du jakin-minak makina, eta hor dago gakoa.

### Bukatzeko

Gure garuna oso konplexua da. Pertzepzioa, etorkizuna auresateko gaitasuna eta jakin-mina garunaren paisaiako elementu bakan batzuk besterik ez dira. Baina adimen artifizialaren esparruan aurrerapen garrantzitsuak eragiteko gai izan dira jada. Badirudi ideia erakargarri horien atzean haiteman daitekeela egiaren traza. Beraz, ikertzen jarraitu beharko dugu hain boteretsuak diruditen ideia horiek behar bezala ulertzeko, eta soluzio berririk bilatu ahal izateko. Epe motzeko eta luzeko memoria, ikasketa interaktiboa ala ikasketa kooperatiboa jada irekita dauden ikerketa lerroak dira [5]. Lanean jarraitu beharko, jakin-minak bultzatzen gaitu eta. ●

### Bibliografia

- [1] Clark, A. (2015). "Surfing uncertainty: Prediction, action, and the embodied mind". *Oxford University Press*.
- [2] Ha, D., & Schmidhuber, J. (2018). "Recurrent world models facilitate policy evolution". *Advances in Neural Information Processing Systems* (pp. 2455-2467).
- [3] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Lillicrap, T. (2018). "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". *Science*, 362(6419), 1140-1144.
- [4] Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). "Large-scale study of curiosity-driven learning". *arXiv preprint arXiv:1808.04355*.
- [5] Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P. A., Strouse, D. J., ... & de Freitas, N. (2018). "Intrinsic Social Motivation via Causal Influence in Multi-Agent RL". *arXiv preprint arXiv:1810.08647*.