

Legile lui Asimov – norme metodologice

MIRCEA BĂDUȚ

Scriitor și inginer

e-mail:

mirceabadut@yahoo.com

Article history

Received: 19.10.2020

Received in revised form: 13.11.2020

Accepted: 15.12.2020

Available online: 30.12.2020

Title: “Asimov’s Laws – methodological norms”

Abstract: The first concerns about the social problems of living with robots arose with the term robot (1920), and in 1942 the writer Isaac Asimov defined the famous 'robot laws', disseminated worldwide through his renowned science-fiction proses. In the following lines we will try to analyze the respective laws seen from the perspective of current and especially future applicability in human society.

Keywords: literature; science-fiction; robots; autonomy; responsibility; morality; ethics; laws; norms;

Preambul

„Ascultând de dimineață la radio – într-o scurtă secvență pe tema roboțiilor și inteligenței artificiale – enunțul celebrelor legi ale lui Asimov, pe loc mi-am zis că *au încurcat-o cei ce-or avea de scris normele metodologice de aplicare!*

Desigur că odinioară am fost fascinat de poveștile pe care Isaac Asimov le-a brodat pe infrastructura legilor roboticii, devenite nu doar legendare ci și, iată!, referință pentru preocupările umanității în perspectiva avansului automaticii și informaticii (și în perspectiva neliniștitoare a *Singularității*¹).

Însă pe atunci nu știam că o lege este un enunț succint, sintetic, și că – pentru a funcționa în practica socială, administrativă, economică și judiciară – ea trebuie deseori completată cu detalii de aplicare concretă, detalii numite 'norme metodologice'.

¹Supoziția teoretică a iminenței unui viitor în care inteligența artificială va fuziona cu inteligența oamenilor, sau chiar o va depăși, preluând eventual controlul lumii.

Legea 1: "Un robot nu are voie să pricinuiască vreun rău unei ființe umane, sau, prin neintervenție, să permită ca unei ființe omenești să i se facă un rău."

Legea 2: "Un robot trebuie să se supună ordinelor date de către o ființă umană, atât timp cât ele nu intră în contradicție cu Legea 1."

Legea 3: "Un robot trebuie să-și protejeze propria existență, atât timp cât acest lucru nu intră în contradicție cu Legea 1 sau Legea 2."

Așa că, pentru cititorii acestor rânduri, aduc provocarea de a reflecta la (și chiar de a imagina) conținutul eventualelor „*norme metodologice pentru legile roboticii*”. Indiferent că acestea ar fi pentru uzul juriștilor (legiuitori, tribunale, judecătorii) sau pentru uzul entităților implicate (constructori și programatori de roboți; proprietari de viitori roboți; roboți conștienți și responsabili legal; ș.a.m.d.)².

Intermezzo

Dintr-o perspectivă esențializată putem admite că morala constă din *regulile de conviețuire* (sau „regulile sociale”, dacă doriți, deși prima sintagmă este mai sugestivă pentru tema de față). Din perspectiva formării individului în societate se poate spune că MORALA ajunge la noi în trei tranșe: (1) prin reflexe biologice intra-specie (instincte primare de socializare, așa cum vedem și la animale); (2) prin educație (prin exemplul oferit de părinți și de cei din jur, și respectiv prin învățarea explicită); (3) prin legile scrise (definite concret de responsabilitii societății). La acest din urmă nivel ne vom referi în cele ce urmează, dar vom privi lucrurile în perspectiva 'inteligenței artificiale' care se presupune că animă roboții.

În căutarea normelor

Înainte de toate se cuvine să recunoaștem că – din perspectiva eventualului conflict dintre oameni și roboți, sau, mai bine zis, dintre oameni și *tehnologia autonomă* (și vă propun această sintagmă alternativă și cuprinzătoare) – formulările realizate de Isaac Asimov sunt admirabile dacă le raportăm la anul emiterii lor: 1942. Adică la doar două decenii după ce Karel Čapek lansase termenul de „robot” în scrierile sale ficționale^[3].

Însă astăzi, un asemenea demers legislativ sintetic ne-ar apărea mai degrabă ca fiind unul pseudo-etic, sau chiar ludic. Textul acelor legi ale roboticii pare desuet, iar din perspectiva aplicabilității ele sunt de-a dreptul perimate. Pe de altă parte, o reformulare la fel de sintetică și cu impact comparabil este puțin probabilă acum. Mentalul societății s-a schimbat prea mult de atunci.

² Fragment din Mircea Băduț, *DonQuijotisme AntropoLexice*, ediția a II-a (București: Editura Europress, București, 2017).

În ultimul timp, am fost martorii unor emanații populare de inteligență artificială (vedeți aplicațiile web/cloud Google Maps și, mai ales, Google Translate) și am putut înțelege un pic din ceea ce înseamnă „machine learning”, ca premisă pentru o viitoare/posibilă autonomie, o *autonomie epistemică* care în anul 1942 nu prea putea fi anticipată. Însă aceasta este doar o parte a punctului de vedere schimbat.

Vă propun, înarmați cu experiența de viață a unor oameni obișnuiți din secolul 21, – deci nu neapărat din perspectiva jurisprudenței – să disecăm un pic textele celor trei legi originare ale roboticii (și poate chiar să le privim din perspectiva unor *eventuale norme metodologice*, așa cum provocasem prin ideea emisă în urmă cu trei ani).

«Legea 1: "Un robot nu are voie să pricinuiască vreun rău unei ființe umane"»

Această primă și esențială parte a legii asimoviene ni se arată docilă din perspectiva aplicării, aspect datorat similitudinii cu clasicele legi ale moralei umane, pentru care există atât cutumă cât și norme scrise în dreptul civil și în cel penal. Lăsăm în urmă sugestia excluziunii din acest enunț (adică speculația aceea afurisită că „*da, robotul nu poate face rău vreunui om, însă ar cam fi liber să facă rău unui alt robot*”), și observăm – extrapolând ideea similitudinii cu legile oamenilor – că ne putem întreba o serie de chestiuni de genul:

- Oare pentru robotul antropomorf (*antropomorf* la propriu dar și la figurat, adică robotul destinat să conviețuiască cu oamenii) nu se aplică în primul rând legile oamenilor, *in integrum*, la care Asimov doar a formulat un „codicil”? Altfel spus, n-ar trebui cumva să considerăm că legile roboticii funcționează ca un *super-set* legislativ, destinat să completeze obligatoriu legile din dreptul civil?

- Cât de autonom și de responsabil (moral și civil) poate fi un robot fabricat și programat de alții? În ce măsură răspunderea legală a faptelor/actelor robotului se împarte cu oamenii sau cu roboții care l-au creat? Mai mult, se poate incrimina penal un algoritm complex, la care participarea creatorilor – oameni sau roboți – a fost una foarte dispersă? (Sau:) Câtă dispersie poate suporta responsabilitatea până să devină... caducă?

Am văzut că deocamdată răspunderea civilă în privința incidentelor infracționale sau contravenționale provocate de automatele existente (precum automobilele autonome Google sau Tesla) este considerată ca aparținând creatorilor și/sau proprietarilor. (Iar dacă este vorba doar de pagube materiale (deci evaluabile), acestea se pot acoperi prin folosirea sistemului de asigurări.) Însă lucrurile se complică în situațiile în care respectivii roboți ajung să evolueze în contexte/circumstanțe neprevăzute, care nu mai pot fi imputate creatorilor sau proprietarilor.

- Care/unde este limita dintre automatul evoluând autonom, deplin responsabil civil, și cel incapabil de discernământ moral? Cum îi numim pe cei care

nu sunt deplin maturați legalmente? Roboți cu răspundere limitată? Androizi minori?

Revenim la textul primei legi a lui Asimov, și anume la partea a doua a enunțului: "*sau, prin neintervenție, să permită ca unei ființe umane să i se facă un rău*". Aici lucrurile sunt de-a dreptul îndoielnice. Da, o normă metodologică poate repara exprimarea laconică, clarificând faptul că se referă la un robot care este *martor al unei vătămări*. (În paranteză vom observa că perspectiva lui Asimov este una incompletă juridic: el se referă doar la infracțiunile având ca obiect direct ființa umană, ignorând efectiv mulțimea de fapte ce pot aduce prejudicii omului și în mod indirect: furt, calomnie, speculă, corupție, minciună, sperjur, fraudă, poluare etc.) Dar chiar și asumând clarificarea eventualei norme de aplicare, tot ne mai rămân aspecte discutabile:

(1) un robot evoluat, cu conexiuni multiple/potente la rețeaua informatică (informațională, de senzori, camere de supraveghere video), ar putea fi teoretic martor la infracțiuni pe o zonă cu acoperire geo-spațială mult mai mare decât cea specifică omului, ceea ce l-ar putea aduce ușor într-o stare de saturație, de inoperare judiciară;

(2) în legea oamenilor nu există o astfel de obligativitate de a interveni într-o infracțiune în desfășurare, așa încât să cerem acest lucru roboților s-ar putea dovedi „incorect politic”. De altfel, aici transpare perspectiva de „sclav al omului” asociată robotului la mijlocul secolului trecut, viziune pusă explicit în text de:

Legea 2. "Un robot trebuie să se supună ordinelor date de către o ființă umană."

Majoritatea oamenilor își imaginează roboții – fie ei industriali, domestici, funcționari de ghișeu, aplicații software, jucării, asistenți medicali, roboți companioni, ș.a.m.d. – ca fiind destinați să slujească oamenilor, pentru că ei chiar sunt automate construite pentru acest scop. Însă în viitor, când/dacă autonomia lor se va extinde – prin sporirea capacităților de memorare, procesare și comunicare – perspectiva s-ar putea schimba. Deja există multe cercetări tehnico-științifice și aplicații practice care dovedesc că inserarea abilităților de autodezvoltare (adăugarea de independență) poate fi o cale de rezolvare a unor probleme mai dificile sau mai complexe. (Eventual facem aici o paralelă epistemologică cu trecerea de la calculatorul von Neumann la cel cuantic.) Iar autodezvoltarea poate însemna atât (1) *acumularea de noi cunoștințe* (creșterea bazei de date prin auto-învățare) cât și (2) *modificarea/optimizarea algoritmilor* de procesare a informațiilor și de confecționare a deciziilor (ceea ce ne aduce iarăși semnul de întrebare privind răspunderea legală). (Deschidem aici o altă paranteză, pentru a observa că în programarea software modernă, începând de la OOP încoace, deja granița dintre date și algoritm nu mai este una strictă. Iar în viitor paradigma s-ar putea deplasa și mai mult.) În privința creșterii autonomiei, pe lângă modelul *machine learning* (ML) amintit anterior, avem în circulație și alte

concepte operaționale: machine-to-machine (M2M), Internet-of-things (IoT), rețele neurale, inteligență artificială (IA). Însă trebuie spus că astfel de sintagme și acronime alcătuiesc adesea o modă frivolă (catalizată de setea pentru *hype* a contemporanității), o emfază ce dovedește speranță, dar ascunde și naivitate și incultură. Iar adesea vehiculează și angoasă (o angoasă nejustificată deocamdată): faptul că avem o mulțime de automate care știu ML, M2M, AI și IoT nu înseamnă deloc că ele se vor autodezvolta în curând până la a se a se „înțărca” singure, pentru a provoca acea *Singularitate* de care civilizația umană se teme.

Despre Legea 3: "Un robot trebuie să-și protejeze propria existență."

Deși în *Carta drepturilor fundamentale a Uniunii Europene* este prevăzut că „Orice persoană are dreptul la integritate fizică și psihică”³, mai nicăieri nu scrie că sinuciderea este ilegală. Altfel zis, pentru oameni propria existență este un drept, nu o obligație. De ce ar fi altfel pentru roboți? Oare pentru că sunt bunuri materiale și nu e bine să le pierdem? Adică un punct de vedere strict economic? Dar mai găsim aici un aspect îndoielnic: pentru a aplica această lege ar trebui ca respectivii roboți să fie conștienți de ei înșiși (fie prin programarea inițială, fie prin autodezvoltare). Apoi, ce înseamnă „conștient de sine”? Și aici putem identifica cel puțin trei niveluri: (1) Cunoștințele memorate (sau senzorii proprii) să informeze robotul asupra gabaritului și/sau asupra gradelor de libertate ale mișcărilor. (2) Conștiința: cunoașterea reflexivă și asumată a propriilor abilități de interacțiune și de modificare a lumii înconjurătoare. (3) Sensul/intuiția unicității, și eventual intuiția perisabilității⁴. De altfel, aceste trei nivele ale conștiinței de sine – fiecare putând corespunde unor nivele de răspundere civilă/legală definibile și respectiv unor niveluri mai mult sau mai puțin implementabile prin algoritmi – pot fi întâlnite și la animale, de la cele multe și simple (mici ierbivore sau carnivore) la cele evaluate psihic (precum elefanții, primatelor sau delfinii).

Concluzii

Încheiem seria de întrebări și dileme făcând o observație oarecum transgresivă: în materie de legislație, civilizația umană are o experiență de cel puțin două mii de ani, deci putem presupune că dificultatea nu rezidă în a defini norme. Adevărata încercare va consta în a defini ce este robotul.

³Vezi Art. 3 (1), [https://eur-](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:ro:PDF)

[lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:ro:PDF](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:ro:PDF)

⁴ Perisabilitatea robotului evoluat poate însemna atât conștientizarea vulnerabilității fizice, cât și finitudinea sa în timp (mortalitatea sa, ca atribut uman). Și ne amintim, tot de la Isaac Asimov, două scrieri foarte reprezentative pentru aceasta „Eu, robotul” și „Omul pozitronic” (în *The Bicentennial Man*, Ballantine Books, 1976).

Bibliografie/References:

Asimov, Isaac. *The Bicentennial Man*. Ballantine Books, 1976.

Băduț, Mircea. *DonQuijotisme AntropoLexice*. București : Editura Europress, 2017.

Čapek, Karel, R.U.R (*Rossum's Universal Robots*). 1920

Carta drepturilor fundamentale a Uniunii Europene. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:ro:PDF>

AUTHOR BIO:

MIRCEA BĂDUȚ is a Romanian writer and engineer. He graduated from the Faculty of Electrical Engineering, University of Craiova, in 1992. He wrote eleven technical-scientific books (of applied informatics) and six books of fiction (short fictional prose; essayistic prose of philosophical, psychological and anthropological nuance). He has written over four hundred technical-scientific articles for magazines in Romania and in Europe. Some of his short prose and of his essays have appeared in various publications and anthologies in Romania.