

L'IMAGINARI SIMBÒLIC DE LA IA

Ciència-ficció: un mirall per al futur de la humanitat

L'interessant debat sobre IA i ètica que propicia la ciència-ficció

Carme Torras



[Araya Peralta](#)

Les tecnologies digitals han esdevingut part de les nostres vides quotidianes i cada cop actuen més com a intermediaris en els nostres llocs de treball i en les relacions personals —fins i tot substituint-les. L'Internet de les coses, les xarxes socials, els programes capaços d'aprendre interactuant amb humans, els assistents artificials i els robots de companyia, els jocs d'ordinador, els jocs amb impacte social, els assessors financers semiautomatitzats, els webs que ofereixen la immortalitat digital, etc. Aquestes eines poden, en un breu lapse de temps, modificar el mercat de treball, acabar amb la reputació de qualsevol, transformar una barriada urbana, canviar les nostres relacions personals —no només a la feina, també dins de la família i en les relacions més íntimes— o magnificar el que un deixa enrere quan es mor, que ara inclou l'anomenada empremta digital.

La creixent interacció amb màquines *intel·ligents* no és sols un pas més enllà en la transformació social que va arrencar amb la revolució industrial. Tot i que aquestes noves tecnologies de la informació també alliberen els humans de tasques repetitives i els proporcionen més temps per invertir-lo de formes creatives i divertides, la diferència és que ara s'endinsen en àmbits que sempre s'han considerat exclusius dels humans, com ara la presa de decisions, les emocions i les relacions socials, i això pot comprometre els valors

humans al mateix temps que, clarament, dona forma a la societat i a la nostra manera de viure.

L'estat de coses actual, sobre el qual no tenim experiència prèvia, genera una sèrie de preguntes ètiques que no van ser pertinents per a altres tipus de màquines del passat; tampoc podem predir amb fiabilitat com acabarà influint aquest nou marc en l'evolució de la humanitat. Això ha portat a la confluència de la intel·ligència artificial (IA) amb les Humanitats, en un debat ètic que està començant a donar els primers fruits, no només amb l'establiment de regulacions i estàndards, sinó també amb iniciatives en àmbits com la docència universitària, la millora professional i la conformació de l'opinió pública. Curiosament, la ciència-ficció sovint té una funció predictiva notable, destacant els pros i els contres dels possibles escenaris del futur. En paraules del famós escriptor Neal Stephenson [1]: «la bona ciència-ficció ofereix una descripció plausible i ben estudiada d'una realitat alternativa on irromp una atractiva innovació tecnològica. L'univers de la bona ciència-ficció es caracteritza per una coherència i una lògica internes que resulten convincents per a científics i enginyers».

Confluència entre la IA i les Humanitats

La comunitat de recerca en IA ha sentit la necessitat de treballar conjuntament amb científics socials, psicòlegs, advocats, filòsofs i antropòlegs per tal d'analitzar les implicacions socials i ètiques de les tecnologies que s'estan desenvolupant i la forma com s'estan desplegant. S'han creat equips multidisciplinars i s'ha donat llum verda a projectes conjunts d'investigació. Les administracions i institucions han volgut prendre-li el pols a l'opinió pública i s'han establert comitès amb l'objectiu que continguin la situació i que generin directrius legals i estàndards comuns.

Considerant això últim, diverses organitzacions i societats professionals estan desenvolupant regulacions per als programadors en IA, dissenyadors en robòtica, empreses i usuaris. L'abril de 2019, el Grup d'Experts d'Alt Nivell sobre IA de la Comissió Europea va presentar l'informe Directrius ètiques per a una IA fiable i, aquell mateix any, l'Associació d'Estàndards IEEE va publicar l'informe Disseny alineat èticament: prioritzar el benestar humà amb sistemes autònoms i intel·ligents, que havia estat sotmès a debat públic des del desembre de 2017. El resultat de la consulta van ser unes tres-centes pàgines amb continguts molt rellevants que posteriorment van revisar un miler d'experts d'àmbits molt diversos, des dels negocis, passant per la universitat, fins a les asseguradores. Anteriorment hi havia hagut altres iniciatives pioneres més modestes, com ara la Declaració de Barcelona per al desenvolupament i l'ús adequats de la IA a Europa i la Declaració de Montréal per un Desenvolupament Responsable de la Intel·ligència Artificial, totes dues durant el 2017.

La regulació legal és necessària per desenvolupar una intel·ligència artificial ètica, però encara és més important l'educació a tots els nivells, amb la finalitat de desenvolupar consciència crítica i formar l'opinió pública. Un objectiu crucial és el col·lectiu de dissenyadors de ginys d'IA. No és sorprenent que en l'àmbit universitari, veus conegudes

com la de la professora Barbara J. Grosz defensin integrar l'ètica en els plans d'estudi d'informàtica. «Fer del raonament ètic un element central dels estudis portarà l'alumnat a aprendre a pensar no només sobre la tecnologia que poden dissenyar, sinó a plantejar-se si haurien de dissenyar-la» [2]. En aquesta línia, associacions prestigioses de la talla de l'ACM i l'IEEE inclouen 18 àrees del saber en els seus currículums d'informàtica, entre els quals hi ha *Aspectes socials i pràctica professional*, per tal que «l'alumnat desenvolupi l'hàbit de pensar sobre els aspectes socials, ètics, legals i professionals pertinents».

A Europa, la Universitat d'Oxford fins i tot ha anat més enllà en la confluència amb les Humanitats, i ofereix un títol en Informàtica i Filosofia, centrat en els interessos comuns entre la intel·ligència artificial, la lògica, la robòtica, la realitat virtual i l'ètica. Aquest tipus de graus mixtos aviat proliferaran arreu; en concret, a Catalunya diverses universitats ja tenen iniciatives que van en aquesta direcció.

La funció de la ciència-ficció

La ciència-ficció ha estat una font d'inspiració per als científics i tecnòlegs des del seu naixement. Treballs clàssics com els de Shelley, Verne i Huxley, entre molts altres, ja concebien experiments, recerques, descobriments i innovacions que més tard han esdevingut realitat, com l'exploració espacial, les rèpliques humanes i l'enginyeria genètica. Avui dia, la ciència-ficció s'ha guanyat l'atenció pública arran de les tecnologies digitals; sobretot la IA i la robòtica. Els darrers anys han estat habituals els contactes entre escriptors i directors de cinema de ciència-ficció, d'una banda, i científics i desenvolupadors de tecnologia, de l'altra.

Com a exemple, el 2012 es va establir el Centre per a la Ciència i la Imaginació a la Universitat Estatal d'Arizona. La idea va sorgir arran el provocador discurs que l'escriptor Neal Stephenson [3] va fer en presència del rector, en el qual afirmava que avui els científics han perdut la capacitat de pensar i de fer «coses grans», com les que en el passat havien inspirat el programa espacial Apollo o el microprocessador. El rector va respondre que potser els responsables de la situació eren els escriptors de ciència-ficció, que havien deixat de suggerir futurs engrescadors que estimulessin els científics a voler fer-los realitat. El resultat va ser el naixement del Centre, que inclou equips d'investigadors en ciències i humanitats treballant conjuntament en projectes que modelaran el nostre futur.

Un dels projectes del Centre és la continuació d'una iniciativa de la companyia Intel, el projecte *Tomorrow*, en què es va demanar a quatre escriptors que escrivissin relats de ciència-ficció sobre els usos possibles que en el futur es podrien donar als productes de la companyia en fotònica, robòtica, telemàtica i sensors intel·ligents. El llibre és d'accés obert [4]. Des de llavors, i fruit del treball del Centre, han aparegut diversos volums amb propostes de solució per als majors reptes als quals la humanitat ha de fer front avui, explorats a través de l'escriptura i les arts visuals.

Són innumbrables les iniciatives sorgides en aquesta direcció: tallers, simposis, exposicions, performances...; fins i tot revistes científiques de gran tirada han publicat

dossiers especials dedicats a la producció en ciència-ficció. Una pionera va ser la prestigiosa revista *Nature*, que per a commemorar el cinquantè aniversari de la hipòtesi de Hugh Everett III sobre els universos paral·lels va publicat un volum titulat *Many Worlds* [5], amb articles tant d'investigadors en mecànica quàntica com d'escriptors de ciència-ficció. La introducció ja deia molt clarament quina funció pot tenir la ciència-ficció enumerant els avantatges i els inconvenients del desenvolupament científic. «La ciència-ficció seriosa es pren la ciència seriosament. [...] La ciència-ficció no ens diu el que ens durà el futur, però sí que ens ajuda a fer-nos-en una idea i ens planteja com podríem sentir-nos quan la manera habitual que tenim de mirar el món sigui substituïda per una altra».

Ensenyar i fomentar el debat ètic a través dels relats de ciència-ficció

Ensenyar ètica en IA dista considerablement d'ensenyar altres matèries en Ciències de la Informació i Enginyeria. No és qüestió que l'alumnat s'apregui uns continguts concrets, sinó de fer-los conscients de les implicacions socials i ètiques de la seva futura feina i d'entrenar-los a analitzar i debatre sobre aquestes qüestions. Els alumnes tenen múltiples que sovint son discrepants, i l'objectiu no és que acabin pensant el mateix, sinó conscienciar-los perquè cultivin la capacitat de reflexionar i debatre. A més, els alumnes de tecnologia no són filòsofs. Tot i que hi ha teories ètiques de les quals haurien de tenir nocions bàsiques, els textos filosòfics són sovint massa abstractes, i dona més bon resultat adoptar una via pragmàtica.

Segons Sullins [6], les teories ètiques pertinents per a la IA i la interacció robot - humà són: conseqüencialisme o utilitarisme (maximitzar el nombre de persones que gaudeixen dels màxims beneficis), deontologisme (actuar només segons màximes que podrien esdevenir lleis universals), ètica de la virtut (confiar en el caràcter moral d'individus virtuoses), justícia social (tot els éssers humans mereixen ser tractats igual; hi ha d'haver una justificació sòlida per al tractament diferencial), el bé comú (viure en comunitat implica certes restriccions per a l'individu), ètica religiosa (les normes deriven d'una autoritat espiritual), i ètica de la informació (reglaments i codis que regeixen la creació, l'organització, la divulgació i l'ús de la informació).

Com que cap teoria per ella mateixa pot donar resposta a tots els aspectes ètics que es deriven del disseny i l'ús de les innovacions tècniques, l'opció més pragmàtica és adoptar una aproximació híbrida. Aquesta ètica híbrida és defensada per Wallach i Allen [7] com a combinació de teories que van de dalt a baix (és a dir, les que fan servir principis racionals per derivar-ne les normes ètiques) i les que van de baix a dalt (les que infereixen les normes generals a partir de situacions concretes).

Ara bé, on trobem aquestes situacions concretes? Stephenson [8] afirma que «allò que les narracions de ciència-ficció fan més bé que cap altra cosa és donar no sols una idea d'innovació tecnològica concreta, sinó també oferir un marc coherent de com aquella innovació serà integrada en la societat, l'economia i la vida de la gent». Així, alguns cursos sobre ètica i tecnologia recorren als relats de ciència-ficció per exemplificar situacions

concretes que visibilitzin el conflicte ètic. Els temes explorats en les obres clàssiques d'Asimov, Dick, Bradbury, Orwell, Huxley, Hoffman, Shelley, Capek, Wells, Sturgeon, Silverberg, o Keyes, com ara les tres lleis de la robòtica, les mainaderes robòtiques, el debat de seguretat versus llibertat, la pèrdua de la privacitat, el totalitarisme tecnològic, els substituïts emocionals, les rèpliques humanes, la incidència en el mercat laboral, la responsabilitat moral, la pèrdua de control per part dels humans, els biaixos de l'alta tecnologia, la distància entre la manipulació i l'automatització o el millorament humà i el posthumanisme són de gran rellevància arran el desenvolupament de la IA.

Donada aquesta rellevància, és natural que els docents d'ètica en els graus tecnològics recorrin a aquest tipus de relats per exemplificar situacions a les quals l'alumnat pot ser sensible o a les quals s'hauran d'afrontar en la pràctica professional, per tal que hi reflexionin o debatin. Després de cinc edicions del curs *Ciència-ficció i ètica computacional* a la Universitat de Kentucky i dues a la Universitat d'Illinois, Chicago, Burton *et al.* [9] afirmen que, en la seva experiència, «recórrer a la ficció per ensenyar ètica permet l'alumnat parlar i raonar amb seguretat sobre qüestions difícils i emocionalment connotades sense portar el debat al terreny personal». A més, destaquen que els relats són engrescadors per als estudiants i fan referència a descobertes molt interessants fetes al llarg dels anys que mereixen una atenta lectura.

La ciència-ficció contemporània toca moltes de les qüestions ètiques de les novel·les clàssiques, però enfocades des d'una perspectiva actual: per exemple, les preocupacions d'avui arran de les tecnologies digitals, l'ús intensiu dels telèfons mòbils, l'estès ús de les xarxes socials, la presa automàtica de decisions guiada per ginys d'intel·ligència artificial, els jocs d'immersió en realitats virtuals, els algoritmes que fan ús del big data... És per això que es generen debats interessants [10]. La figura 1 mostra algunes sèries i pel·lícules centrades en aquestes qüestions. Quant a això, la sèrie *Black Mirror* és una obra mestra que, en cada capítol, porta una tecnologia particular a les seves conseqüències extremes, i la pel·lícula *Her* retrata un home que s'enamora del sistema operatiu del seu ordinador, traslladant el títol de Hoffmann *The Sandman* a un paradigma digital contemporani. Pel que fa a la robòtica, destacaria la sèrie *Real Humans*, on robots quasi-humans coexisteixen amb humans i sovint competeixen amb ells, i la pel·lícula *Surrogates*, on cada ciutadà té un avatar controlat des de casa que fa moure per la ciutat i que interactua amb humans. La pel·lícula *Robot and Frank* mostra la relació entre un Frank envellit i el seu cuidador artificial, i mereix especial esment pel seu realisme i valor educatiu: fou la base per a un curs en línia sobre ètica i robòtica a través de la plataforma 'Teach with Movies' [11], entre altres.



Figura 1. Sèries i pel·lícules de ciència-ficció que tracten qüestions ètiques generades per les tecnologies digitals, especialment la IA i la robòtica.

Pel que fa a llibres pensats per a l'ensenyament de cursos d'ètica en graus tecnològics a través de la ciència-ficció, ressaltaria els de Murphy [12], Nourbakhsh [13] i Torras [14]; vegeu la figura 2. L'editat per Murphy explica els principis clau de la IA; els relats que inclou —d'Asimov, Vinge, Aldiss i Dick— toquen les qüestions de la telepresència, la robòtica especialitzada en la imitació del comportament, la deliberació, la validació, la interacció robot-humà, la vall inquietant, el processament de llengües naturals, l'aprenentatge automàtic i l'ètica. Cada narració és precedida per una nota introductòria, 'Mentre llegeixes', i va seguida d'una discussió sobre les possibles derivades, 'Després de la lectura'.

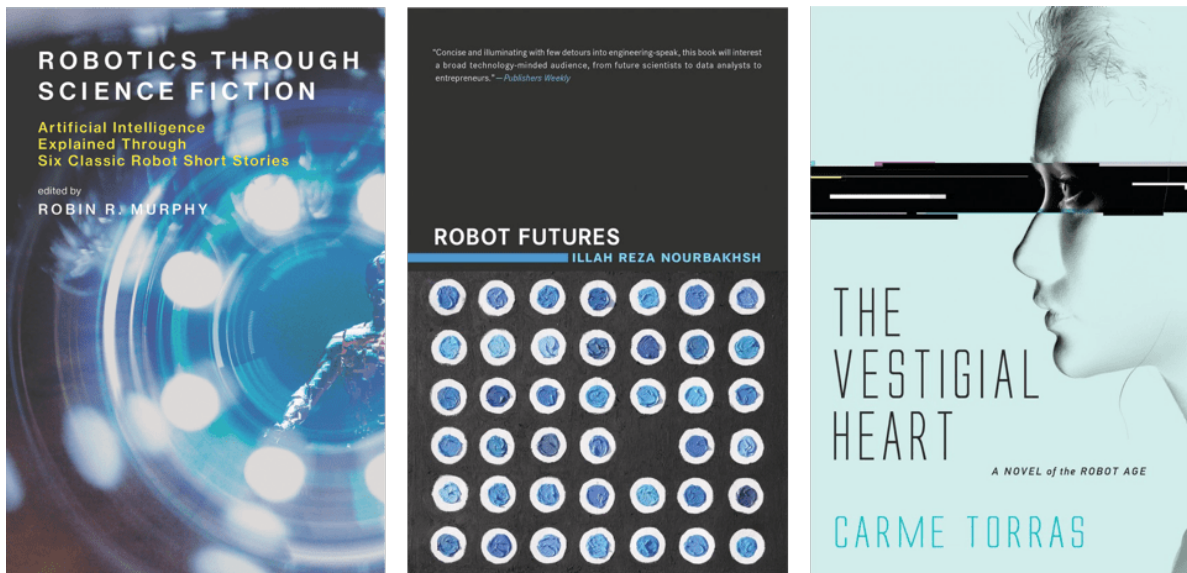


Figura 2. Tres llibres publicats per MIT Press especialment pensats per a estudiants tecnològics i basats en relats de ciència-ficció.

Els altres dos llibres als quals em referiré difereixen del primer en dos aspectes: d'una banda, els autors són també autors dels relats de ciència-ficció utilitzats com a il·lustració i, de l'altra, els llibres no inclouen explicacions sobre principis tècnics, però busquen explícitament fomentar el debat sobre les implicacions socials i ètiques de les tecnologies de la informació i la comunicació.

Amb la descripció de possibles escenaris de futur, Nourbakhsh genera preguntes sobre cap on és que estem avançant, i ho fa sense adoptar ni un punt de vista regulador ni un to marcadament pedagògic. Com a famós investigador en robòtica, l'autor fa comentaris molt lúcids sobre temes com les estratègies de màrqueting a la xarxa; conseqüències del disseny no efímer; joguines robòtiques voladores que operen seguint la mirada; telepresència multimodal i multicontinental habilitada per mitjà d'un robot, i fins i tot la possibilitat d'adoptar diferents formes físiques gràcies als nanorobots. Nourbakhsh examina la tecnologia en què cada giny està basat i descriu les conseqüències socials de cada escenari. Però també ofereix una rèplica: una robòtica dissenyada per a enfortir els llaços cívics i comunitaris.

En virtut de la meua recerca en assistents artificials [15], em vaig interessar progressivament per les implicacions socials i ètiques de les tecnologies que estem desenvolupant, concretament per trobar maneres d'ensenyar ètica als estudiants amb perfils tecnològics. Això em dugué a provar la meua habilitat per escriure ficció: en la novel·la *La mutació sentimental* vaig explorar com el fet de ser criat per mainaderes artificials, o de tenir mestres robòtics i de compartir feina i lleure amb programes d'IA, afectaria els hàbits intel·lectuals, emocionals i socials de les generacions futures. El *leitmotiv* de la novel·la és una cita del filòsof Robert C. Solomon: «són les relacions que hem construït les que al seu torn ens modelen» [16]. Ell es referia a les relacions humanes amb els nostres pares, mestres i amics, però la frase es pot aplicar també als assistents artificials i a tota mena de dispositius interactius, si és que han de ser tan presents en les nostres vides.

La trama de la novel·la és la següent: la Cèlia, una noia de tretze anys congelada criogenèticament a causa d'una malaltia terminal, és curada i portada de nou a la vida en ple segle XXII per ser adoptada. A part dels records, tragina alguna cosa més del passat: sensacions i emocions que ja no existeixen en el nou món. Això és el que més atreu la Silvana, una dona de mitjana edat que treballa com a massatgista emocional mirant de recuperar les sensacions que els humans han perdut. Els sentiments de la Cèlia són també un preuat material de recerca per a en Leo, un bioenginyer que està dissenyant una pròtesi de creativitat per al misteriós Dr. Craft, propietari de la important companyia de robòtica CraftER.

Seguint un suggeriment de l'editora de MIT Press Marie L. Lee, vaig incloure en el llibre un apartat final amb 24 preguntes de contingut ètic i apunts per generar debat sobre les situacions que apareixen en la novel·la. Es va publicar conjuntament amb una guia en línia per al professor i una presentació de 100 diapositives per impartir un curs sobre ètica en robòtica social i intel·ligència artificial. El llibre cobreix sis temes: com dissenyar l'«assistent» perfecte; la importància de l'aparença i de la simulació d'emocions en els robots per a ser acceptats; la funció dels programes d'IA en el lloc de treball i en l'entorn educatiu; el dilema entre la presa de decisions automàtica i la llibertat i dignitat humanes, i la responsabilitat implícita d'allotjar cert sentit moral en els programes informàtics i els robots.

Cada capítol de la guia del professor segueix la mateixa estructura: arrenca amb alguna secció destacada de la novel·la per passar a introduir després el marc ètic de la situació conflictiva, seguida de quatre preguntes i punts per estimular el debat. El capítol es tanca amb algunes qüestions ja tractades en capítols anteriors. En la figura 3 apareixen algunes diapositives de mostra de l'estructura d'aquests materials docents.

Teaching materials
Ethics in Social Robotics and AI

Appendix in "*The Vestigial Heart*":

1. Designing the "perfect" assistant: Chapters 1, 5
2. Robot **appearance** and emotion: Chapters 9,10,12
3. Robots in the **workplace**: Chapter 13
4. Robots in **education**: Chapters 14, 16
5. Human-robot **interaction** and human dignity: Chapters 25, 28
6. **Social responsibility** and robot morality: Chapter 30



Teaching materials
Ethics in Social Robotics and AI

5. Human-robot interaction and human dignity

- 5.1. Highlights from *The Vestigial Heart*
- 5.2. Ethical Background and Discussion:
 - Four questions
 - Hints for a debate on each question
- 5.3. Revisiting Issues
- 5.4. Scholarly References for Further Reading

Designing the "perfect" assistant
1.2. Questions



- 1.A - Should public trust and confidence in AI/robots be enforced? If so, how?
- 1.B - Is it admissible that devices be designed to generate addiction?
- 1.C - Should the possibility of deception be actively excluded at design time?
- 1.D - Could AI be used to control people?

"Ethics in Social Robotics" based on *The Vestigial Heart* @ MIT Press, 2018 18/104

Human-robot interaction and human dignity
5.2. Questions

- 5.A - Could automatic decision-making undermine human freedom and dignity?
- 5.B - Is it acceptable for robots to behave as emotional surrogates? If so, in what cases?
- 5.C - Could robots be used as therapists for the mentally disabled?
- 5.D - How adaptive/tunable should programs and robots be? Are there limits to human enhancement by technology?

"Ethics in Social Robotics" based on *The Vestigial Heart* @ MIT Press, 2018 66/104

Figura 3. Mostra de diapositives dels materials docents associats amb el llibre *La mutació sentimental*: a sobre, correspondència entre les seccions en què es distribueix el contingut ètic i els capítols de la novel·la, i estructura de cada secció; a sota, algunes de les preguntes plantejades.

El llibre i els materials de suport docent han estat utilitzats no només en els graus de Ciències de la Informació i Enginyeria, sinó també en els estudis de Filosofia i Administració d'empreses de diverses universitats del món; de moment, principalment a Espanya i als EUA, però algunes universitats europees ja han mostrat interès en poder comptar-hi en els propers cursos. A més, hi ha una versió catalana dels materials adaptada a la secundària i [disponible al web de Pagès Editors](#), i que també pot servir per a generar debat en grups de lectura, panells, taules rodones i altres fòrums similars de discussió.

A tall de conclusió

La creixent interacció dels humans amb programes d'IA en tota mena de contextos quotidians genera importants reptes socials i ètics que molt probablement modificaran el nostre futur. Això exigeix que els graus tècnics s'acostin a les Humanitats, de manera que alumnes i professionals es facin conscients dels possibles aspectes que hauran d'afrontar al llarg de la seva carrera i aprenguin a definir-los i a debatre sobre ells.

Hem descrit tres propostes recents d'investigadors de l'àmbit de la IA i la robòtica: un llibre que utilitza relats clàssics per explicar els principis computacionals i dos llibres que confien en la ciència-ficció contemporània per transmetre als alumnes el saber tecnoètic, amb la característica singular que les ficcions que contenen són també obra dels mateixos

investigadors. Així, els relats il·lustren exactament els aspectes que els autors volen tractar, i el material és pedagògicament engresgador tant per a lectors amb perfil tecnològic com per als docents i probablement també per a qualsevol lector aficionat a llegir sobre alternatives de futur.

Conclouré dient que la major part de la literatura de ciència-ficció és respectuosa amb la ciència i mira de projectar les fites científiques cap al futur. Sembla que la ciència comença ara a respectar també aquesta literatura visionària i a trobar-hi cert objecte d'inspiració. La confluència entre els dos àmbits podria ser summament productiva i és una molt bona notícia, atès que obre perspectives interessants per al futur immediat.

RECONeixEMENTS

Aquest treball ha estat parcialment possible gràcies al suport del Consell de Recerca Europeu, sota el programa de recerca i innovació Horitzó 2020 de la Unió Europea, a través del projecte CLOTHILDE - CLOTH manipulation Learning from DEMonstrations (acord de concessió núm. 741930), i l'Agència de Recerca Espanyola a través del Segell d'Excel·lència Maria de Maeztu concedit a l'IRI (MDM-2016-0656).

REFERÈNCIES

- 1 — Stephenson, N. (2011) "Innovation starvation". *World Policy Journal*, 28(3), 11-16. Disponible en línia a:
<http://www.worldpolicy.org/journal/fall2011/innovation-starvation>
- 2 — Grosz, B.J.; Grant, D.G.; Vredenburg, K.; Behrends, J.; Hu, L.; Simmons, A. i Waldo, J. (2019) "Embedded EthICS: integrating ethics across CS education". *Communications of the ACM*, 62(8), 54-61.
- 3 — Stephenson, N. (2011) "Innovation starvation". *World Policy Journal*, 28(3), 11-16. Disponible en línia a:
<http://www.worldpolicy.org/journal/fall2011/innovation-starvation>
- 4 — Rushkoff, D.; Hammond, R.; Thomas, S. i Markus, H. (2012) *The Tomorrow Project*. Els millors escriptors descriuen la vida quotidiana en el món del futur. Intel: Santa Clara.
- 5 — Everett III, H. (2007) «Many Worlds» *Nature*, 448(7149): 1-104.
- 6 — Sullins, J.P. (2015) "Applied professional ethics for the reluctant roboticist". Dins de *The Emerging Policy and Ethics of Human-Robot Interaction*, editat per L.D. Riek, W. Hartzog, D. Howard, A. Moon i R. Calo. Taller de treball dins el 10è Congrés Internacional de l'ACM/IEEE sobre la Interacció Humans-Robots, Portland.
- 7 — Wallach W. i Allen C. (2008) *Moral machines: Teaching robots right from wrong*. Oxford: Oxford University Press.
- 8 — Stephenson, N. (2011) "Innovation starvation". *World Policy Journal*, 28(3), 11-16. Disponible en línia a:
<http://www.worldpolicy.org/journal/fall2011/innovation-starvation>
- 9 — Burton, E.; Goldsmith, J. i Mattei, N. (2018) "How to teach computer ethics through science fiction". *Communications of the ACM*, 61(8), 54-64.

- 10 —
- Torras C. (2018) "Social networks and robot companions: Technology, ethics and science fiction". *Metode Science Studies Journal*, 99: 47-53.
 - Torras C. i López de Mántaras, R. (ed.) (2019) "Interlinked: machines and humans facing the 10101 century". *Metode Science Studies Journal - Annual Review*, 9.
- 11 — Teach with Movies (2012) "Robot Ethics Using Clips from *Robot and Frank*". <http://teachwithmovies.org/robot-and-frank/>
- 12 — Murphy, R. (ed.) (2018) *Robotics Through Science Fiction: Artificial Intelligence Explained Through Six Classic Robot Short Stories*. MIT Press: Cambridge, Massachusetts.
- 13 — Nourbakhsh, I.R. (2013) *Robot Futures*. MIT Press: Cambridge, Massachusetts.
- 14 — Torras C. (2018) *The Vestigial Heart. A Novel of the Robot Age*. MIT Press: Cambridge, Massachusetts (llibre del professor i presentació de 100 diapositives com a material de suport per al docent (en anglès) disponible en línia a: <http://mitpress.mit.edu/books/vestigial-heart>)
- 15 —
- Torras C. (2016) "Service robots for citizens of the future". *European Review*, 24(1), 17-30.
 - Torras C. (2019) "Assistive robotics: Research challenges and ethics education initiatives". *DILEMATA: International Journal of Applied Ethics*, 30: 63-77.
- 16 — Solomon, R.C. (1977) *The Passions*. Nova York: Anchor Press / Doubleday.



Carme Torras

Carme Torras és novel·lista i professora investigadora a l'Institut de Robòtica i Informàtica Industrial, un centre de recerca conjunt del Consejo Superior de Investigaciones Científicas (CSIC) i la Universitat Politècnica de Catalunya (UPC). És llicenciada en Matemàtiques i Informàtica per la Universitat de Barcelona i la Universitat de Massachusetts, respectivament, i té un doctorat en informàtica per la UPC. És IEEE Fellow, EurAI Fellow, integrant de l'Acadèmia Europea i membre de la Reial Acadèmia de Ciències i Arts de Barcelona. Compagina la direcció d'un grup de recerca en robòtica assistencial amb la seva faceta d'escriptora literària, i promou l'ètica en l'aplicació de les noves tecnologies. És autora de diversos llibres i articles, i també ha escrit novel·les de ciència-ficció relacionades amb la robòtica i la intel·ligència artificial.