

ESTEM PERDENT EL CONTROL?

Edward A. Lee

Universitat de Califòrnia, Berkeley, EUA. eal@berkeley.edu

Traducció: Aina Barceló Cuerda, Societat Catalana de Tecnologia

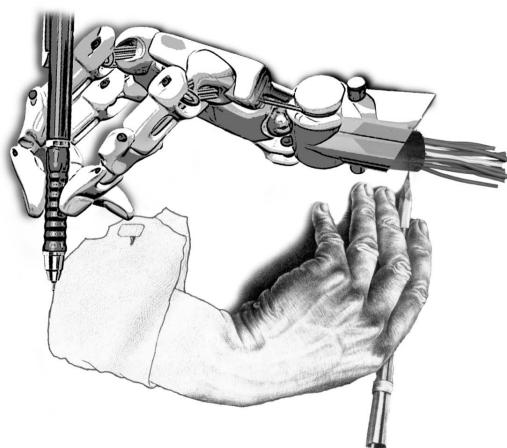


FIGURA 1. La tecnologia digital coevoluciona amb els éssers humans.
FONT: Imatge cedida per l'autor.

Context de la traducció

Recentment s'ha publicat en accés obert el llibre *Perspectives en humanisme digital* (Hannes Werthner, Erich Prem,

Edward A. Lee i Carlo Ghezzi (ed.), *Perspectives on digital humanism*,¹ Springer, 2022), que ha estat possible gràcies a la iniciativa sobre humanisme digital que s'inicià l'any 2019 a Viena i que agrupa persones d'arreu del món i diferents disciplines per parlar de la relació entre les tecnologies de la informàtica i la societat. A la Societat Catalana de Tecnologia sentim que aquest tema ens toca de prop; és, sens dubte, una gran satisfacció publicar aquí la traducció al català del primer capítol d'aquest llibre,² fet que posa en relleu el nostre afany de divulgació de l'esperit crític envers la tecnologia a la nostra societat.

Edward A. Lee fa quaranta anys que treballa en sistemes de programari incrustat i, després de passar per Yale, el Massachusetts Institute of Technology (MIT) i Bell Labs, va aterrar a Berkeley, on ara és professor de l'Escola de Postgrau en Enginyeria Elèctrica i Informàtica. La seva recerca se centra en els sistemes ciberfísics, àmbit en què s'esforça per fer sistemes sensibles i alhora componibles, segurs i verificables. Recentment s'ha diversificat i ha publicat dos llibres sobre filosofia de la tecnologia, *Plató and the nerd* (2017) i *The coevolution* (2020).

Resum: Aquest assaig qüestiona la suposició predominant que els humans donen forma a la tecnologia mitjançant un disseny descendent intel·ligent, suggerint que la tecnologia hauria de ser vista com el resultat d'un procés evolutiu darwinian en què els humans són els agents de mutació. En conseqüència, els humans tenim molt menys control del que pensem sobre els resultats del desenvolupament tecnològic.

Paraules clau: tecnologia, societat, intel·ligència artificial.

ARE WE LOSING CONTROL?

Resum: This essay challenges the predominant assumption that humans shape technology using top-down, intelligent design, suggesting that technology should instead be viewed as the result of a Darwinian evolutionary process where humans are the agents of mutation. Consequently, we humans have much less control than we think over the outcomes of technology development.

Keywords: technology, society, artificial intelligence.

1. <https://dighum.ec.tuwien.ac.at/perspectives-on-digital-humanism/>.

2. Edward A. LEE, «Are we losing control?», a Hannes WERTHER, Erich PREM, Edward A. LEE i Carlo GHEZZI (ed.), *Perspectives on digital humanism*, Springer, 2022. Aquest capítol està publicat sota una llicència Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Estem perdent el control?

El canvi ràpid genera por. Temem que la intel·ligència artificial (IA), que en l'última dècada hem vist renéixer de les seves cendres, pugui substituir la majoria dels treballadors de coll blanc (Ford, 2015); aprengui a millorar iterativament fins a convertir-se en una superintel·ligència que deixaria la humanitat enrere (Barrat, 2013; Bostrom, 2014; Tegmark, 2017); fragmenti la informació de tal manera que les persones quedin dividides i isolades en conjunts disjunts de veritats (Lee, 2020); substitueixi la presa de decisions humanes en l'atenció sanitària, les finances i la política (Kelly, 2016); enforteixi els poders autoritaris, rastrejant cada moviment de la ciutadania i modelant els seus pensaments (Lee, 2018), i que els monopolis capitalistes de vigilància, que necessiten la IA, destrueixin les petites empreses i ofeguin l'emprenedoria (Zuboff, 2019).

Realment, avui dia encara conservem una espurna de control. Com a mínim, encara podem desconnectar l'endoll. N'estem segurs? La tecnologia subjacent a aquests riscos la fan humans; llavors, per què no podem controlar-ne les conseqüències? Tenim el poder de dissenyar i de regular, oi? Aleshores, per què estem intentant tan desesperadament solucionar els desastres d'ahir, mentre que el d'avui tot just acaba de produir-se? La mateixa tecnologia que ens amenaça també fa possible alimentar amb èxit la majoria dels 7.800 milions d'humans d'aquest menut planeta i que milers de milions de persones hagin sortit de la pobresa en les últimes dècades. Reflexionem, no obstant, amb la famosa frase d'Albert Einstein: «No podem resoldre els nostres problemes amb el mateix pensament que vam fer servir quan els vam crear».

El coneixement és a l'arrel de la tecnologia, la informació és a l'arrel del coneixement i la tecnologia actual fa que la informació sigui molt més accessible del que mai ha estat. Això no ens hauria d'ajudar a resoldre els nostres problemes? L'explosió de la IA alimenta el tsunami: converteix cada imatge, cada text i cada so en més informació, i inunda els nostres febles cervells humans. No podem absorbir el flux sense sotmetre'l prèviament a una cura de continguts, i la cura de la informació la fan cada cop més les IA. Cada subconjunt de la veritat és només una veritat parcial, i la informació seleccionada inclou, necessàriament, un subconjunt. Com que el nostre cervell només pot absorbir un petit subconjunt del flux, tot el que absorbim és, en el millor dels casos, una veritat parcial. Les IA, en canvi, semblen tenir poques dificultats amb la riuada. Per a elles, és l'aliment que les enforteix, potser conduint a aquest temut bucle de superintel·ligència que deixa al marge els humans, en la irrellevància.

La pregunta que abordo aquí és: «Nosaltres estem perdent el control?». Potser trobareu inquietant la meva resposta.

En primer lloc, cal plantejar aquesta pregunta, què entenem per «nosaltres»? És la «humanitat», els 7.800 milions d'humans? La idea que 7.800 milions de persones controlin qualsevol cosa col·lectivament és evidentment

absurda, així que no ha de ser això el que volem dir. Ens referim als enginyers de Silicon Valley? Als inversors de Wall Street? Als polítics que ens proporcionen veritats parcials i mentides descarades?

En segon lloc, què entenem per «control»? Seria com conduir un cotxe per una xarxa de carreteres, o més aviat conduir-lo mentre el mapa es va creant i transformant en carrerons sense sortida, passos inferiors i bucles inesperats? Si estem conduint la tecnologia, aleshores cada gir que fem canvia el terreny al qual ens dirigim de manera inesperada.

Soc enginyer. A la meua manera, contribueixo al problema escrivint programari, una part del qual influeix en el nostre ecosistema. Durant bona part dels meus quaranta anys fent això he anat albergant la il·lusió «creacionista» que les coses que dissenyava eren la meua creació, el resultat pur de les meves decisions deliberades, la meua pròpia producció creativa. M'he adonat que això seria una mica com pensar que la bossa de queviures que porto del supermercat és el meu èxit personal. Ignoraria segles de desenvolupament de la tecnologia agrícola responsable de la increïble varietat d'aliments frescos de la botiga, del sistema econòmic que permet que tot això sigui assequible, de la tecnologia del cotxe amb què he arribat fins aquí, i de molts altres aspectes socioculturals de fons que fan empal·lidir el meu modest èxit.

Al meu llibre recent (Lee, 2020), empro el terme *creacionisme digital* per a la idea que la tecnologia és el resultat d'un disseny intel·ligent descendent. Aquest principi suposaria que cada tecnologia és el resultat d'un procés deliberat, on cada aspecte d'un disseny és el resultat d'una decisió humana intencionada. Ara sé, quaranta anys després, que no és així com passa. Els programadors serien més aviat els agents de la mutació en un procés evolutiu darwinianà. El resultat dels seus esforços està més condicionat pels ordinadors, xarxes, programari, biblioteques, llenguatges de programació i altres programes que no pas per les seves decisions deliberades. I l'èxit i el desenvolupament posterior del seu producte estan determinats tant o més per l'entorn cultural al qual llancen la seva «creació» que per les seves decisions de disseny.

El filòsof francès conegut com a *Alain* (el nom real del qual era Émile-Auguste Chartier) va escriure sobre els vaixells de pesca a Bretanya:

Cada vaixell es copia d'un altre vaixell... Raonem sobre això com Darwin. És evident que un vaixell molt mal fet anirà al fons després d'un o dos viatges, i així no es copiarà mai... Per tant, podem dir, en sentit estricte, que és el mateix mar el que modela els vaixells, escollint els adequats i destruint els altres. (Rogers i Ehrlich, 2008)

Els dissenyadors de vaixells són agents de la mutació i, de vegades, les seves mutacions donen lloc a un vaixell mal fet. Sota aquesta perspectiva, potser el disseny de Facebook ha estat més modelat pels adolescents que pels programadors.

Més profundament, la tecnologia digital coevoluciona amb els humans. Facebook canvia els seus usuaris, que després canvien Facebook. Per als programadors, les eines que fem servir, creades prèviament per la mateixa enginyeria de programari, configuren el nostre pensament. Penseu en com els IDE³ (com Eclipse o Visual Studio Code), els taulers de missatges (com ara Stack Overflow), les biblioteques (com la Standard Template Library), els llenguatges de programació (Scala, Rust i JavaScript, per exemple) i la cerca a Internet (com Google o Bing) afecten el resultat del nostre programari. Aquestes eines tenen més efecte en el resultat que totes les nostres decisions deliberades.

Avui en dia, la por i l'exageració al voltant de la IA que s'apodera del món i les xarxes socials i enderroca la democràcia alimenten l'exigència de més regulació. Però, si tinc raó sobre la coevolució, potser ens estem equivocant a l'hora de regular la tecnologia. Per què les lleis de privadesa, amb totes les seves bones intencions, han aconseguit poc per protegir la nostra privadesa? Només ens han aclaparat amb textos legals en lletra petita i finestres emergents molestes que ens donen la possibilitat d'escollir entre «acceptar els nostres termes inescrutables» i «anar-se'n». Creiem que seran més efectives les noves regulacions que intentin mitigar notícies falses o evitar insurreccions instigades per xarxes socials?

Sota el principi del creacionisme digital, els mals resultats són el resultat d'accions poc ètiques per part dels individus, per exemple, seguint cegament l'ànim de lucre sense preocupar-se pels efectes socials. Sota el principi de coevolució, els mals resultats són el resultat de la «proesa procreativa» (Dennett, 2017) de la mateixa tecnologia. Les tecnologies que tenen èxit són les que es propaguen amb més eficàcia. Els individus als quals atribuïm l'origen (o en culpem) d'aquestes tecnologies, sens dubte hi tenen un paper, però també ho fan els usuaris de les tecnologies i tot el seu context cultural. Sota aquesta perspectiva, els usuaris de Facebook tenen part de la culpa, juntament amb Mark Zuckerberg, dels resultats distorsionats a les eleccions. Fins i tot tenen part de la culpa del disseny del programari de Facebook que ho permet. Si acceptéssim xarxes socials de pagament, per exemple, podríem tenir un disseny de programari completament diferent.

Sota el creacionisme digital, el propòsit de la regulació és limitar els individus que desenvolupen i comercialitzen tecnologia. En canvi, sota la coevolució, les limitacions poden ser sobre l'ús de la tecnologia, no només el seu disseny i comercialització. El propòsit de la regulació esdevé impulsar el procés tant de la tecnologia com de l'evolució cultural mitjançant incentius i sancions. L'impuls és probablement el millor que podem esperar. Els processos evolutius no cedeixen fàcilment el control perquè naveguem en un terreny de canvi continu.

3. Un entorn integrat de desenvolupament (IDE, en anglès) és una eina informàtica per ajudar els programadors analitzant el seu text mentre escriuen, acolorint text per funció, identificant errors i possibles defectes en l'estil del codi, suggerint insercions, i en la refacció del codi.

Potser les lleis de privadesa han estat ineficaces perquè es basen en el creacionisme digital com a principi. Aquestes lleis assumeixen que canviar el comportament de corporacions i enginyers serà suficient per assolir els objectius de confidencialitat (siguin els que siguin per a cadascú). Una perspectiva coevolucionària entén que els usuaris de la tecnologia optaran per renunciar a la confidencialitat encara que se'ls digui explícitament que es farà un ús abusiu de la seva informació. Exactament això apareix reiteradament a la lletra petita de totes aquelles polítiques de confidencialitat que no llegim, i, tanmateix, els nostres fills es veuen inmersos en un entorn mediàtic on la seva identitat queda definida clarament com a no privada.

Si la tecnologia determina la cultura mentre que la cultura determina la tecnologia, tenim un bucle de retroalimentació, i la intervenció en qualsevol punt del bucle de retroalimentació pot canviar els resultats. Per tant, podria ser tan efectiu aprovar lleis centrades en l'educació pública, per exemple, com ho seria aprovar lleis que regulen els productors de tecnologia. Potser si més gent entengués que Pokémon GO és un motor de modificació del comportament, entendria millor la política de privadesa de Niantic i la seva afirmació que el seu producte, Pokémon GO, no té publicitat. Els negocis paguen a Niantic per col·locar un Pokémon a prop per atraure la gent a visitar-los (Zuboff, 2019). Potser també hauria de formar part del remei enfortir les lleis de difamació, contra el discurs d'odi i altres millores dels drets de la primera esmena.⁴

Crec que, com a societat, podem fer-ho millor del que estem fent actualment. El risc d'un estat orwellià (o potser pitjor, un gran germà corporatiu) és molt real. Ja ha passat a la Xina. No ho farem millor, però, fins que abandonem el creacionisme digital com a principi. Prohibir desenvolupaments tecnològics específics no serà efectiu, i trencar els monopolis podria empitjorar el problema en accelerar les mutacions. Per exemple, podem intentar prohibir la presa de decisions autònoma en els sistemes d'armes i la banca, però, com veiem a partir de les distorsions electorals i Pokémon GO, les IA són molt efectives per influir en la presa de decisions humanes, de manera que posar un ésser humà en el bucle no ajuda necessàriament. Com pot una persona que, realment, està controlada per una màquina, d'alguna manera mitigar la maldat de les armes autònomes?

Quan parlo d'educar el públic, molta gent immediatament es precipita cap a una solució màgica, on la bala de plata seria d'ensenyar ètica als enginyers. Permeteu-me la pregunta: si suposem que tots els tecnòlegs es comporten de manera ètica (sigui quin sigui el significat d'aquesta paraula), podem concloure que no es produiran mals resultats? Això em sembla ingenu. Els processos coevolutius són massa complexos.

Aquest assaig és la meua modesta contribució a la iniciativa d'humanisme digital, un moviment que busca un

4. Esmena 1 de la Constitució dels Estats Units, referida a la llibertat de religió, d'expressió, de premsa, de reunió i de petició (n. d. t).

enfocament més humà en la tecnologia. Aquesta iniciativa fa imprescindible que intel·lectuals de totes les disciplines facin un pas endavant i es prenguin seriosament la dansa de la humanitat amb la tecnologia. El fet que els nostres esforços limitats per frenar els efectes perjudicials de la tecnologia digital hagin estat majoritàriament ineficaços subratlla la nostra feble comprensió del problema. Necessitem humanistes amb una comprensió més profunda de la tecnologia, tecnòlegs amb una comprensió més profunda de les humanitats i responsables polítics d'ambdós camps. Avui estem força lluny d'aquest objectiu.

Tornant a la pregunta original, estem perdent el control? La resposta és «no». Mai no hem tingut el control, i no podem perdre allò que no tenim. Això no vol dir que ens hem de rendir, però. Podem desviar el procés, igual que també un vaixell superpetrolier es podria redirigir només amb una dòcil empenta.

Bibliografia

BARRAT, J. (2013). *Our final invention: Artificial intelligence and the end of the human era*. Nova York: St. Martin's Press.

BOSTROM, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford, Regne Unit: Oxford University Press.

DENNETT, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. Nova York: W. W. Norton and Company.

FORD, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Nova York: Basic Books.

KELLY, K. (2016). *The inevitable: Understanding the 12 technological forces that will shape our future*. Nova York: Penguin Books.

LEE, E. A. (2020). *The coevolution: The entwined futures of humans and machines*. Cambridge, Massachusetts: MIT Press.

LEE, K.-F. (2018). *Super-Powers: China, Silicon Valley, and the new world order*. Nova York: Houghton Mifflin Harcourt Publishing Company.

ROGERS, D. S.; EHRLICH, P. R. (2008). «Natural selection and cultural rates of change». *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, núm. 9, p. 3416-3420.

TEGMARK, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Nova York: Alfred A. Knopf.

ZUBOFF, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Nova York: Public-Affairs: Hachette Book Group.