

Chiquita Room

MODELS DE MIRAR

UNA EXPOSICIÓ D'ESTAMPA

14.11.2024 - 21.12.2024

QUÈ VOL DIR VEURE UNA PERSONA?

El problema de la visió ha atret cada cop més l'atenció de la crítica a l'era de la visió artificial. L'algorisme YOLO9000, explorat per Estampa amb pinzellades lúdiques en aquesta exposició, il·lustra perfectament la naturalesa d'aquesta qüestió. Yolo és un sistema de detecció d'objectes conegut per la seva precisió i velocitat. El nom li ve donat per una promesa que, en realitat, podria ser una amenaça: «You Only Look Once». A la seva cèlebre sèrie de televisió i llibre de 1972, *Ways of Seeing*, el crític d'art John Berger posa el focus en que veure mai no ha estat només un procés fisiològic, sinó que també implica hàbits i convencions culturals¹. Tant per a les màquines com per als humans —que cada cop depenen més de grans i petites màquines per formar-se una imatge del món— veure implica emmarcar. La fotografia ha perfeccionat la tècnica de col·locar rectangles al món mòbil en un intent d'estabilitzar-lo. Al llarg dels anys, la teoria de la fotografia ha desenvolupat un compromís intencional amb aquesta pràctica. Escriptors i escriptors com Susan Sontag, John Tagg, Ariella Aïsha Azoulay i Mark Sealy² han argumentat de diverses maneres que, encara que les càmeres capturen objectes i els mantenen per ajudar-nos a donar sentit al món, aquest procés és només temporal i parcial. Si bé hi ha certa violència en aquests actes de captura, podem minimitzar-la enfrontant-nos al nostre desig de congelar, controlar i posseir el món.

Tanmateix, aquesta consciència crítica al voltant de les imatges no sembla haver-se transferit a la informàtica convencional. El seu subcamp estrella actual, la visió artificial, es basa en una sèrie de truc conceptuals i tècnics en la seva implementació de la «visió»: la supressió de la profunditat dins la imatge, la fusió del primer pla amb el fons i l'aplanament dels objectes en patrons. D'aquesta manera, els objectes tridimensionals del món real es representen com a models bidimensionals mitjançant la selecció de les seves «característiques». La imatge se segmenta inserint-hi múltiples marcs interns per aïllar algunes entitats. Com a resultat, una fotografia es redueix a «una col·lecció d'objectes per etiquetar» en comptes de ser vista com «un tot cohesionat o una composició amb significat relacional»³. La relacionalitat en la fotografia s'estén al context

¹ Vegeu John Berger (1972) *Ways of Seeing*. Londres: British Broadcasting Corporation and Penguin Books.

² Vegeu Susan Sontag (1977) *On Photography*. Nova York: Farrar, Straus and Giroux; John Tagg (1988) *The Burden of Representation: Essays on Photographies and Histories*. Amherst: University of Massachusetts Press; Ariella Aïsha Azoulay (2019) *Potential History: Unlearning Imperialism*. Londres: Verso, 2019; Mark Sealy (2019) *Decolonising the Camera: Photography in Racial Time*. Londres: Lawrence & Wishart.

³ Amanda Wasielewski (2023) «Authenticity and the Poor Image in the Age of Deep Learning», *photographies* 16(2): 191-210, 195-196. Per a més informació sobre aquest tema, vegeu Joanna Zylińska (2023) *The Perception Machine: Our Photographic Future Between the Eye and AI*. Cambridge: MIT Press; Joanna Zylińska (2024) «Diffused Seeing: The Epistemological Challenge of Generative AI», *Media Theory*, publicació especial a «Photographic Seeing», Vol. 8 N.º 1: 229-258.

històric i social, la construcció deliberativa i, sovint, conflictiva del significat a través de categories i conceptes, que sempre són més que etiquetes. (Podem provar d'associar representacions imaginatives amitjanades a conceptes com «veritat», «democràcia» i «justícia», per exemple, però inevitablement es produiran algunes generalitzacions tosques, per no parlar de la comissió d'actes de violència epistèmica, atès que es plantejaran preguntes respecte les decisions preses sobre les imatges, les formes d'exclusió iniciades i la imatge del món perpetuada a través d'aquest acte mateix de representar i etiquetar).

Per descomptat, els informàtics són conscients de les acusacions de «biaix» adreçades als seus resultats. Tot i això, la majoria de les respostes a aquest problema han estat tècniques, basades en la creença que un conjunt de dades més ampli permetria posar fi a qualsevol biaix evident en les dades. Fa més d'una dècada, el grup de recerca liderat pel professor Fei-Fei Li de la Universitat de Stanford va donar una nova arrencada a la visió artificial en expandir la base de dades de fotografies ImageNet de milers a milions d'artefactes, extraient-los de diferents parts d'internet. La clau del seu èxit va raure en fer servir la fàbrica clandestina d'Amazon, MTurk, perquè treballadors anònims etiquetessin aquests milions d'imatges i els parells imatge-text resultants es van emprar a gran escala per entrenar models en el reconeixement d'objectes; actualment un procés molt ràpid que Estampa mostra de manera commovedora. En els darrers anys hem estat testimonis de diversos intents per anar més enllà de la planitud de les imatges en la visió artificial que elideix el gruix fenomenològic del món físic. I així, el 2024, Fei-Fei Li ha recaptat 230 milions de dòlars per a l'*start-up* World Labs, que té per objectiu crear tecnologia d'IA que «pot entendre com funciona el món físic tridimensional»⁴. El projecte promet llançar la «intel·ligència espacial» mitjançant el «desenvolupament de [...] "grans models mundials" que faran servir professionals com ara artistes, dissenyadors, desenvolupadors i enginyers»⁵.

Però, aquesta darrera iniciativa planteja algunes preguntes: què entendreà *realment* aquesta tecnologia? Quina mena d'intel·ligència inclourà i promourà; quin tipus d'espais habitarà, i quin tipus de mons voldrà construir? És important assenyalar que el fet de plantejar aquestes preguntes no suposa tornar a l'excepcionalisme humà, suggerint que l'ésser humà és l'única espècie que pot desenvolupar una comprensió veritable i que les màquines només seran la nostra pura imitació. Els darrers avenços en grans models lingüístics han demostrat que, en efecte, aquest no és el cas. Tanmateix, sí suposa reprendre la pregunta que emmarca aquest

⁴ Anna Tong i Katie Paul (2024) «AI godmother' Fei-Fei Li raises \$230 million to launch AI startup», *Reuters*, 13 de setembre, <https://www.reuters.com/technology/artificial-intelligence/ai-godmother-fei-fei-li-raises-230-million-launch-ai-startup-2024-09-13/>.

⁵ Marina Temkin (2024) «Fei-Fei Li's World Labs comes out of stealth with \$230M in funding», *TechCrunch* In Brief, 14 de setembre, <https://techcrunch.com/2024/09/13/fei-fei-lis-world-labs-comes-out-of-stealth-with-230m-in-funding/>.

assaig: *què vol dir veure una persona?* A través del treball d'Estampa, podríem matisar encara més aquesta pregunta: *què vol dir veure una persona que és un treballador?* Com la veiem? Quins espais ocupa? Què hi fa? Amb quin propòsit i en quines condicions? En la seva forma actual, sembla que la instrumentalitat de veure en visió artificial continua sent un punt cec a l'hora d'examinar de veritat aquestes qüestions. El megasforç per veure a través de tots els espais i punts de dades probablement acabarà amitant tots els valors atípics al conjunt de dades, amb la consegüent supressió total del desordre econòmic i ecològic de la imatge.

Joanna Zylinska

Artista, escriptora i professora de Filosofia dels Mitjans i Pràctica Digital Crítica al King's College de Londres. És autora de diversos llibres, entre ells, *The Perception Machine* (2023), *AI Art: Machine Visions and Warped Dreams* (2020) i *Nonhuman Photography* (2017). A la seva pràctica artística experimenta amb diferents tipus de mitjans basats en imatges.

MODELS DE MIRAR

Mirem i ens miren contínuament. El món se'ns apareix a través de les pantalles i desperta en nosaltres el desig de comunicar-nos-hi, de produir i intercanviar imatges amb la nostra xarxa de relacions. L'interès econòmic per extreure dades estadístiques de tot aquest torrent d'imatges socials ha requerit, els últims anys, la construcció de nous enginys de visió. Un dels més estesos a la xarxa és el de l'anotació automàtica de les imatges. A cada nova publicació en línia, s'activen operacions, invisibles als ulls humans, que identifiquen i etiqueten tot allò que reconeixen en les imatges. Aquests processos, adherits a la imatge digital, conformen el revers de la cultura visual, una agència de doble sentit en la qual el qui mira les imatges és també mirat.

El terme «mirada» també el fem servir per referir-nos a una cosmovisió o una cultura. Parlem de teories o autors que ens ofereixen una mirada particular sobre tal temàtica, per exemple. La mirada sempre implica aquesta perspectiva pròpia o un punt de vista específic sobre el món. Aconseguir que els ordinadors «vegin» també suposa la construcció d'aquests sistemes a partir d'una mirada particular, la qual coincideix amb els valors de qui els ha dissenyat. Són eines de visió que codifiquen i posen en circulació determinades concepcions del món en forma de models matemàtics. Són models de mirar. I són portadors d'una mirada que actua per delegació.

Aquests models de mirar estan basats en l'assignació de paraules a fragments significatius de cada imatge. Els guia una pulsio ordenadora que opera per acumulació, formant una teranyina d'etiquetes que pretenen explicar i domesticar les imatges. Però aquest procés no resulta exempt d'alteracions. Quan imbriquem imatge i text, tenim dues opcions: o emprar el text per explicar la imatge, per tancar-ne el significat, o la possibilitat de multiplicar-ne els sentits. El cas és que els models de mirar, dissenyats per funcionar segons la primera opció, sovint acaben concurrent en la segona. Aquest fenomen, que per als enginyers que hi treballen pot ser un maldecap, des del punt de vista creatiu es torna ben interessant, perquè certs descriptors influeixen de manera inesperada en la interpretació de la imatge original, produint una intensificació i plantejant un repte per la imaginació de l'espectador. Aquestes fugues de sentit, les friccions provocades per l'intent de capturar allò representat, ens remetent a la inacabable qüestió de superposar paraules i imatges, a la seva tensió, als límits i possibilitats poètiques d'aquesta relació.

La gran majoria de les imatges aquí exposades no són el producte de manipular la imatge ni de trastejar amb els models d'intel·ligència artificial (IA), són el resultat d'un procés de curadoria d'accidents produïts pels automatismes de la visió artificial. Amb curiositat, ens hem dedicat a processar pel·lícules del nostre arxiu i a posar l'atenció en aquests equívocs i col·leccionar-los amb simples captures de pantalles. En aquest gest, els models són extrets de la seva funció operativa per

mudar en una funció poètica que evidencia part del seu fracàs. En alguns casos, la sortida gràfica de les etiquetes sí que ha estat manipulada, desbordant la imatge original fins al punt de saturar-la. La ubiqüitat de les deteccions, habitualment ocultes a la visió, revela que les imatges avui són hostes d'un llenguatge secret que les habita i les transforma. Cada imatge digital esdevé, d'aquesta manera, una frontera entre la compulsió extractivista i totes les fugues fantasmals de la representació, entre la nostra mirada i la mirada de les imatges.

En el context occidental, concebem la mirada en termes receptius. A partir de l'emmirallament cultural amb els enginyers òptics, entenem que mirar equival a rebre estímuls de l'exterior a través de la vista. Però en altres cultures mirar també pot entendre's en termes projectius. Diu l'antropòleg Roger Canals que en l'àmbit digital la nostra mirada reforça una dimensió més projectiva, pel fet que és una mirada que deixa traça en forma de dades. La simple observació d'imatges implica avui un índex estadístic que serà processat en algun remot centre de dades. Però, i en el sentit contrari? Com afecta la nostra mirada que les imatges ens estiguin mirant? L'estigma de les paraules deixa traça en les nostres retines? En aquest cas, faríem bé en obrir les paraules i les imatges, en desfer la seva normativitat espectral, conjurant relacions inesperades.

Estampa

Col·lectiu artístic de programadors, realitzadors i investigadors amb *residència* a Barcelona. La seva pràctica es basa en una aproximació crítica i arqueològica a les tecnologies audiovisuals i digitals, amb un interès particular per l'arxiu i l'audiovisual experimental. Des del 2017 una de les seves principals línies de treball ha estat la investigació dels usos i ideologies de la IA.

ACTIVITATS

Dijous 14 de novembre, 19 h.

Inauguració de l'exposició

Dijous 21 de novembre, 19 h.

Presentació del llibre *crater lover worker* i projecció de peces audiovisuals d'Estampa:

- ¿Qué es lo que ves, YOLO9000? (2019) – 13:43
- Espècies marcianes (2021) – 02:47
- The Vertigo of the Ways of Seeing (2022) – 17:02

Chiquita Room