

IA generativa

lloc: [Institut d'Ensenyaments a Distància de les Illes
Balears](#)
Curs: Models d'intel·ligència artificial
Llibre: IA generativa

Imprès per: Carlos Sanchez Recio
Data: dimarts, 25 de març 2025, 07:23

Taula de continguts

1. Intel·ligència artificial generativa

2. Història de la IA generativa

2.1. Xarxes neuronals generatives

2.2. Esclat de la IA generativa

3. Modalitats

3.1. Text

3.2. Codi

3.3. Text a imatge

3.4. Àudio i música

3.5. Text a vídeo

3.6. Accions

3.7. Modelització 3D

4. Programari i maquinari

5. Regulació

6. Qüestions ètiques i legals

7. Tothom pot ser artista?

8. Enginyeria de prompt

8.1. Conèixer el model

8.2. Especificitat

8.3. Prompts contextuals

8.4. Exemples

8.5. Prompts i personatges

8.6. Chain of Thought

8.7. Conclusió

9. Casos d'ús

9.1. Ganyet - Mazoni

9.2. Boris Eldagsen

9.3. DreamUp

10. Un univers en expansió

11. Reptes

12. Enllaços

1. Intel·ligència artificial generativa

Trobam la **intel·ligència artificial generativa** amb les expressions generative AI, GenAI o GAI.

Podem definir-la com a IA capaç de generar **textos, imatges, vídeos** o altres dades utilitzant models generatius.

Sovint, aquesta generació es realitza a partir dels anomenats **prompts**, les descripcions de text que l'usuari introdueix als sistemes d'IA generativa.

Els models d'IA generativa aprenen els patrons i l'estructura de les seves dades d'entrenament d'entrada i a partir d'aquí són capaços de generar noves dades que tindran unes característiques semblants.

Les millores en les xarxes neuronals profundes basades en transformers varen permetre l'esclat dels sistemes d'IA generativa dels inicis dels 2020. Aquests sistemes inclouen diverses categories.

Xatbots basats en models de llenguatge extensos ([MLE](#), en anglès LLM, Large Language Models):

- ChatGPT
- Copilot
- Gemini
- LLaMA

Sistemes de generació d'imatge:

- Stable Diffusion
- Midjourney
- DALL-E

Generadors IA de text a vídeo:

- Sora

Entre les empreses que desenvolupen models d'IA generativa destaquen les següents, vora d'altres companyies més petites.

- OpenAI
- Anthropic
- Microsoft
- Google
- Baidu

La IA generativa presenta casos d'ús en un ventall ampli d'indústries, incloent-hi el desenvolupament de programari, l'assistència sanitària, finances, entreteniment, atenció al client, vendes i màrqueting, art, escriptura, moda i disseny de producte.

Tanmateix, també ha sorgit preocupació sobre el mal ús potencial de la IA generativa en el cibercrim, la difusió de notícies falses (*fake news*) o *deepfakes* per enganyar o manipular les persones, així com la possible substitució massiva del treball humà.

2. Història de la IA generativa

Des de l'establiment de la intel·ligència artificial al taller de recerca del Dartmouth College el 1956, la disciplina ha experimentat diverses onades d'avanç i optimisme al llarg de dècades. Des dels inicis, els investigadors han elaborat arguments filosòfics i ètics sobre la naturalesa de la ment humana i les conseqüències de crear éssers artificials amb una intel·ligència comparable a la humana. Aquestes qüestions s'havien explorat abans en els mites, la ficció i la filosofia des de l'antiguitat. El concepte d'art automàtic existeix com a mínim des dels autòmats de la civilització grega clàssica, quan s'atribuí a inventors com Dèdal o Heró d'Alexandria haver dissenyat màquines capaces d'escriure text, generar sons i reproduir música. La tradició dels autòmats creatius a florit al llarg de la història, amb l'exemple destacat de l'autòmat de Maillardet dels inicis del segle XIX.

La idea de la intel·ligència artificial ha captivat la imaginació de la societat des de mitjans del segle XX. Va començar amb la ciència ficció popularitzant la idea però la idea no es va considerar científicament fins que Alan Turing tingué curiositat sobre la viabilitat del concepte. L'any 1950 Turing publicà un article que plantejà qüestions fonamentals sobre un possible raonament matemàtic semblant a la intel·ligència humana, una contribució fonamental a les bases de la IA.

Des de la fundació de la IA la dècada de 1950, els artistes i investigadors han usat la IA per a crear obres artístiques. A l'inici de la dècada de 1970, Harold Cohen feia i exposava obres d'IA generativa creades amb AARON, un programa que Cohen desenvolupà per produir pintures.

Les cadenes de Markov s'han usat de fa tants per modelitzar els llenguatges naturals d'ençà que el matemàtic rus Andrey Markov les desenvolupà a l'inici del segle XX. Markov publicà el seu primer article sobre el tema el 1906, i analitzà el patró de vocals i consonants de la novel·la Eugeny Onegin usant cadenes de Markov. Després que una cadena de Markov estàs entrenada sobre un corpus de text, es pot usar com a generador probabilístic de text.

El camp de l'aprenentatge automàtic sol fer ús de models estadístics, entre ells els models generatius, per modelitzar i predir dades. Començant a finals dels 2000, l'emergència de l'aprenentatge profund impulsà el progrés i recerca en classificació d'imatge, reconeixement de la parla, [processament del llenguatge natural](#) i d'altres tasques. Les xarxes neuronals d'aquesta època se solien entrenar com a models discriminatius, a causa de les dificultats de la modelització generativa.

El 2014, avenços com l'autocodificador variacional (*variational autoencoder*) i les xarxes generatives antagonistes produïren les primeres xarxes neuronals profundes capaces d'aprendre models generatius, en oposició als discriminatius, per a tasques complexes com les imatges. Aquests models profunds generatius varen ser els primers a produir no només etiquetes de classe per a les imatges sinó també imatges senceres.

El 2017, la xarxa transformer permeté avenços en els models generatius en comparació als models LSTM. Això va dur al primer transformer generatiu preentrenat (GPT, Generative Pretrained Transformer), conegut ara com GPT-1, el 2018. El 2019 GPT-2 demostrà la capacitat de generalitzar de forma no supervisada a moltes tasques, com a model fundacional.

El 2021, DALL-E, un model generatiu de píxels basat en transformer, seguit per Midjourney i Stable Diffusion, suposà l'emergència d'art IA d'alta qualitat a partir de prompts de text.

El març de 2023 es va llançar GPT-4, que alguns veien com una versió inicial i incompleta d'intel·ligència artificial general (AGI) i d'altres l'hi consideren encara molt enfora. El 2023, Meta llança un model d'IA anomenat ImageBind que combina dades de text, imatges, vídeo, tèrmiques, 3D, àudio i moviment, del qual s'espera la generació de contingut IA generatiu més immersiu.

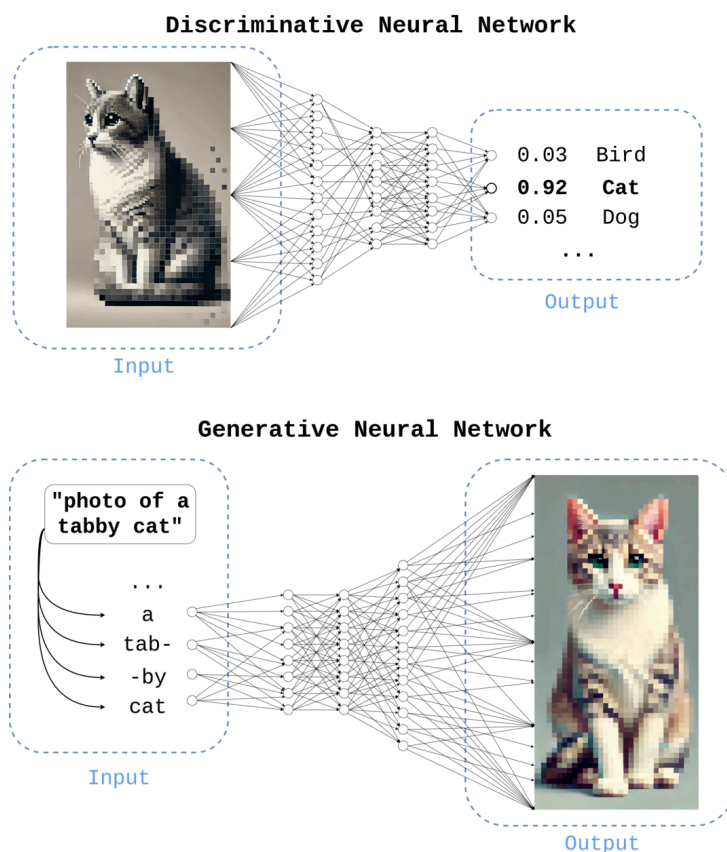
2.1. Xarxes neuronals generatives

Des dels seus inicis, el camp de l'aprenentatge automàtic va utilitzar tant models discriminatius com generatius, per modelitzar i predir dades. A partir de finals dels anys 2000, l'auge de l'aprenentatge profund va impulsar el progrés i la investigació en la classificació d'imatges, el reconeixement de la parla, el [processament del llenguatge natural](#) i altres tasques. Les xarxes neuronals d'aquesta època s'entrenaven normalment com a models discriminatius, a causa de la dificultat de la modelització generativa.

El 2014, avanços com l'**autocodificador variacional** (*variational autoencoder*) i les **xarxes antagonistes generatives** (*generative adversarial network*) donaren les primeres xarxes neuronals profundes pràctiques capaces d'aprendre models generatius, en oposició als models discriminatius, per a dades complexes com per exemple les imatges. Aquests models generatius profunds varen ser els primers que donaven com a sortida no només etiquetes de classe per a les imatges sinó també imatges completes.

El 2017, la xarxa Transformer permeté avanços en els models generatius en comparació amb els models LSTM més antics, que va portar al primer transformer preentrenat generatiu (GPT), conegut ara com a GPT-1, el 2018. Seguidament, el 2019 GPT-2 va demostrar la capacitat de generalitzar a moltes tasques diferents de forma no supervisada.

Els nous models generatius introduïts durant aquest període permetien que les xarxes s'entrenassin usant aprenentatge no supervisat o semisupervisat, en lloc del típic aprenentatge supervisat dels models discriminatius. L'aprenentatge no supervisat eliminà la necessitat que persones etiquetassin les dades manualment, i això permetia l'entrenament de xarxes més grans.



2.2. Esclat de la IA generativa

El març de 2020 es va publicar 15.ai, una aplicació web lliure que podia crear veus de personatges convincents usant unes mínimes dades d'entrenament. Va ser un dels primers casos d'ús populars de la IA generativa. S'atribueix a aquesta plataforma el primer servei que va popularitzar la clonació de veu per IA en memes i creació de continguts, cosa que va influir els desenvolupaments posteriors en tecnologia de veu.

El 2021, l'aparició de DALL-E, un model generatiu de píxels basat en transformer, va marcar un avanç en la imatgeria generada per IA. A continuació varen sorgir Midjourney i Stable Diffusion el 2022, que universalitzaren l'accés a la creació artística a partir d'indicacions en llenguatge natural. Aquests sistemes mostraren unes capacitats sense precedents per generar imatges fotorealistes, pintures i dissenys basats en descripcions de text, i varen ser adoptats massivament per artistes, dissenyadors i el públic en general.

A finals de 2022, el llançament de GPT-4 va ser un salt qualitatiu en les capacitats de la IA generativa. Un equip de Microsoft Research va argumentar de manera controvertida que "podria considerar-se raonablement com una versió primerenca (tot i que encara incompleta) d'un sistema d'intel·ligència artificial general (AGI)". Tanmateix, aquesta avaluació va ser impugnada per altres estudiosos que sostenien que la IA generativa era "encara lluny d'assolir el punt de referència de la intel·ligència humana general". Més tard, el 2023, Meta va llançar ImageBind, un model d'IA que combina múltiples modalitats, com ara text, imatges, vídeo, dades tèrmiques, dades 3D, àudio i moviment, obrint el camí per a aplicacions d'IA generativa més immersives.

El desembre de 2023, Google va presentar Gemini, un model d'IA multimodal disponible en quatre versions: Ultra, Pro, Flash i Nano. La companyia va integrar Gemini Pro al seu chatbot Bard i va anunciar plans per a "Bard Advanced" impulsat pel model Gemini Ultra més gran. El febrer de 2024, Google va unificar Bard i Duet AI sota la marca Gemini, llançant una aplicació mòbil a Android i integrant el servei a l'aplicació Google a iOS.

El març de 2024, Anthropic va llançar la família Claude 3 de grans models lingüístics, incloent Claude 3 Haiku, Sonnet i Opus. Els models van demostrar millores significatives en les capacitats en diversos punts de referència, amb Claude 3 Opus superant notablement els models líders d'OpenAI i Google. El juny de 2024, Anthropic va llançar Claude 3.5 Sonnet, que va demostrar un rendiment millorat en comparació amb el Claude 3 Opus més gran, especialment en àrees com la codificació, els fluxos de treball de diversos passos i l'anàlisi d'imatges.

Segons una enquesta de SAS i Coleman Parkes Research, la Xina s'ha convertit en un líder mundial en l'adopció d'IA generativa, amb el 83% dels enquestats xinesos utilitzant la tecnologia, superant tant la mitjana global del 54% com la taxa dels EUA del 65%. Aquest lideratge s'evidencia encara més amb l'evolució de la propietat intel·lectual de la Xina en aquest camp, amb un informe de l'ONU que revela que les entitats xineses van presentar més de 38.000 patents d'IA generativa entre 2014 i 2023, superant substancialment els Estats Units en sol·licituds de patents.

3. Modalitats

Un sistema d'IA generativa es construeix aplicant aprenentatge automàtic no supervisat o bé autosupervisat a un conjunt de dades. Les capacitats d'aquest sistema dependran de la modalitat o tipus de dades que s'hi han fet servir.

La IA generativa pot ser unimodal o multimodal. Els sistemes unimodals prenen només un tipus d'entrada, mentre que els sistemes multimodals poden prendre més d'un tipus de dades. Per exemple, una versió de GPT-4 d'OpenAi accepta com a entrada text i imatge.

- Text
- Codi
- Imatges
- Audio (i música)
- Vídeo
- Molècules
- [Robòtica](#)
- Planificació
- Dades
- Disseny assistit per ordinador

3.1. Text

Els sistemes d'IA generativa entrenats sobre paraules o tokens de paraules inclouen GPT-3, LaMDA, LLaMA, BLOOM, GPT-4, Gemini i d'altres.

GPT3 és un model de llenguatge extens fet públic per OpenAI el 2020. Com el seu predecessor GPT2, és una xarxa neuronal profunda de tipus transformer només amb decodificador, que supera les arquitectures basades en recurrència i convolució mitjançant la tècnica de l'atenció. Aquest mecanisme d'atenció permet que el model s'enfoqui de forma selectiva en segments del text d'entrada que preveu com a més rellevants. Usa una finestra de 2048 tokens, precisió de 16 bits i 175 milers de milions de paràmetres. Cada paràmetre usa 2 bytes, per tant, necessita 350GB d'emmagatzematge. Ha demostrat habilitats zero-shot i few-shot en moltes tasques.

LaMDA (Language Model for Dialogue Applications) és una família de grans models de llenguatge conversacional desenvolupat per Google. Desenvolupat i introduït originalment com a Meena el 2020, el LaMDA de primera generació es va anunciar durant la conferència de Google I/O del 2021, mentre que la segona generació es va anunciar l'any següent. El juny de 2022, LaMDA va guanyar una atenció generalitzada quan l'enginyer de Google Blake Lemoine va afirmar que el xatbot havia adquirit consciència. La comunitat científica ha rebutjat en gran mesura les afirmacions de Lemoine, tot i que ha provocat converses sobre l'eficàcia de la prova de Turing, que mesura si un ordinador pot passar per un humà. El febrer de 2023, Google va anunciar Bard, un xatbot d'intel·ligència artificial de conversa impulsat per LaMDA, per contrarestar l'auge del ChatGPT d'OpenAI.

LLaMA (Large Language Model Meta AI) és una família de models extensos de llenguatge autoregressius, llançada per Meta AI a partir del febrer de 2023.

Per a la primera versió de LLaMA, es van entrenar quatre mides de model: 7, 13, 33 i 65 milers de milions de paràmetres. Els desenvolupadors de LLaMA van informar que el rendiment del model de paràmetres 13B a la majoria de punts de referència de NLP superava el del GPT-3 molt més gran (amb paràmetres de 175B) i que el model més gran era competitiu amb models d'última generació com PaLM i Chinchilla. Mentre que els LLM més potents generalment només han estat accessibles a través d'API limitades (si n'hi ha), Meta va llançar els pesos del model de LLaMA a la comunitat de recerca sota una llicència no comercial.

El juliol de 2023, Meta va llançar diversos models com Llama 2, utilitzant 7, 13 i 70 milers de milions de paràmetres.

BLOOM (BigScience Large Open-science Open Access Multilingual Language Model) és un model de llenguatge gran autorregressiu (LLM) basat en transformadors de 176 milers de milions de paràmetres. El model, així com el codi base i les dades utilitzades per entrenar-lo, es distribueixen sota llicències lliures. BLOOM es va entrenar amb aproximadament 366 mil milions de tokens (1,6 TB) de març a juliol de 2022.

BLOOM és el principal resultat de la iniciativa col·laborativa BigScience, un taller de recerca d'un any de durada que va tenir lloc entre maig de 2021 i maig de 2022. BigScience va ser dirigit per HuggingFace i va comptar amb la participació de diversos centenars d'investigadors i enginyers de França i de l'estranger representant tant el món acadèmic i el sector privat. BigScience va comptar amb el suport d'una subvenció informàtica pública a gran escala sobre la supercomputadora pública francesa Jean Zay, gestionada per GENCI i IDRIS (CNRS), sobre la qual es va formar.

El corpus d'entrenament de BLOOM, anomenat ROOTS, combina dades extretes de la darrera versió del corpus OSCAR basat en web (38% de ROOTS) i dades recentment recollides extretes d'una llista de fonts de dades d'idiomes seleccionada i documentada manualment. Comprèn 46 llenguatges naturals i 13 llenguatges de programació.

Generative Pre-Trained Transformer 4 (GPT-4) és un gran model de llenguatge multimodal creat per OpenAI i el quart de la seva sèrie de models de base GPT. Es va llançar el 14 de març de 2023 i es va posar a disposició pública mitjançant el producte de chatbot de pagament ChatGPT Plus, l'API d'OpenAI i el chatbot gratuït Microsoft Copilot. Com a model basat en transformadors, GPT-4 utilitza un paradigma on s'utilitza la formació prèvia amb dades públiques i "dades amb llicència de proveïdors de tercers" per predir el següent testimoni. Després d'aquest pas, el model es va ajustar amb retroalimentació d'aprenentatge de reforç dels humans i la IA per a l'alineació humana i el compliment de les polítiques.

Els observadors van informar que la iteració de ChatGPT mitjançant GPT-4 va ser una millora respecte a la iteració anterior basada en GPT-3.5, amb l'advertència que GPT-4 conserva alguns dels problemes de les revisions anteriors. GPT-4, equipat amb capacitats de visió (GPT-4V), és capaç de prendre imatges com a entrada a ChatGPT. OpenAI s'ha negat a revelar diversos detalls tècnics i estadístiques sobre GPT-4, com ara la mida precisa del model.

Aquests sistemes tenen capacitats de [processament del llenguatge natural](#), traducció automàtica i generació de llenguatge natural.

Es poden usar com a models fundacionals per a d'altres tasques.

Els conjunts de dades amb què es poden entrenar inclouen BookCorpus, Wikipedia i d'altres.

3.2. Codi

A més del text en llenguatge natural, es poden entrenar grans models de llenguatge sobre text de llenguatge de programació, cosa que els permet generar codi font per a nous programes informàtics. Per exemple, OpenAI Codex és un model d'intel·ligència artificial desenvolupat per OpenAI. Analitza el llenguatge natural i genera codi com a resposta. És a la base de GitHub Copilot, una eina d'autocompleció de programació per a IDE seleccionats, com Visual Studio Code i Neovim. Codex és un descendent del model GPT-3 d'OpenAI, ajustat per utilitzar-lo en aplicacions de programació.

OpenAI va llançar una API per a Codex en beta tancada. El març de 2023, OpenAI va tancar l'accés a Codex. A causa de les crides públiques dels investigadors, OpenAI va invertir el rumb. Els investigadors del programa d'accés a la recerca OpenAI encara poden utilitzar el model de Codex.

3.3. Text a imatge

Un model de text a imatge és un model d'aprenentatge automàtic que pren una descripció d'entrada en llenguatge natural i produeix una imatge que coincideix amb aquesta descripció.

Els models de text a imatge es van començar a desenvolupar a mitjans de la dècada de 2010 durant els inicis del boom de la IA, com a resultat dels avenços en les xarxes neuronals profundes. El 2022, es va començar a considerar la producció de models de text a imatge d'última generació, com ara DALL-E 2 d'OpenAI, Imagen de Google Brain, Stable Diffusion de Stability AI i Midjourney, per apropar-se a la qualitat de fotografies reals i art dibuixat per humans.

Els models de text a imatge són generalment models de difusió latent, que combinen un model de llenguatge, que transforma el text d'entrada en una representació latent, i un model d'imatge generatiu, que produeix una imatge condicionada a aquesta representació. Els models més eficaços s'han entrenat generalment amb quantitats massives de dades d'imatge i text extretes del web.

Història

Abans de l'auge de l'aprenentatge profund, els intents de construir models de text a imatge es limitaven a collages mitjançant l'ordenació d'imatges de components existents, com ara una base de dades d'imatges clip art.

La tasca inversa, els subtítols d'imatges, era més manejable, i una sèrie de models d'aprenentatge profund de subtítols d'imatges van ser anteriors als primers models de text a imatge.

El primer model modern de text a imatge, alignDRAW, va ser introduït el 2015 per investigadors de la Universitat de Toronto. alignDRAW va ampliar l'arquitectura DRAW introduïda anteriorment (que utilitzava un codificador automàtic variacional recurrent amb un mecanisme d'atenció) per condicionar-se a seqüències de text. Les imatges generades per alignDRAW tenien una resolució petita (32 × 32 píxels, obtinguts a partir del canvi de mida) i es consideraven de "poca diversitat". El model va poder generalitzar-se a objectes no representats a les dades d'entrenament (com ara un autobús escolar vermell) i va gestionar adequadament noves indicacions com ara "un senyal d'aturada vola sobre cel blau", mostrant resultats que no només "memoritzava" les dades del conjunt d'entrenament.

El 2016, Reed, Akata, Yan et al. varen ser els primers a utilitzar xarxes adversàries generatives per a la tasca de text a imatge. Amb models entrenats en conjunts de dades estrets i específics del domini, van poder generar imatges "visualment plausibles" d'ocells i flors a partir de subtítols de text com "un ocell totalment negre amb un bec gruixut i arrodonit". Un model entrenat en el conjunt de dades COCO (Common Objects in Context) més divers va produir imatges que eren "des de la distància... encoratjadores", però que no tenien coherència en els detalls. Els sistemes posteriors inclouen VQGAN-CLIP, XMC-GAN i GauGAN2.

Un dels primers models de text a imatge que va captar l'atenció generalitzada del públic va ser el DALL-E d'OpenAI, un sistema transformador anunciat el gener de 2021. Un successor capaç de generar imatges més complexes i realistes, DALL-E 2, es va presentar l'abril de 2022, seguit de Stable Diffusion que es va publicar l'agost de 2022. La personalització de text a imatge permet ensenyar al model un nou concepte utilitzant un petit conjunt d'imatges d'un objecte nou que no s'incloua en el conjunt d'entrenament del model de base text a imatge. Això s'aconsegueix mitjançant la inversió textual, és a dir, la recerca d'un nou terme de text que correspongui a aquestes imatges.

Seguint altres models de text a imatge, les plataformes de text a vídeo basades en models de llenguatge com Runway, Make-A-Video, Imagen Video, Midjourney i Phenaki poden generar vídeo a partir de missatges de text i/o text/imatge.

Arquitectura i entrenament

Els models de text a imatge s'han creat utilitzant una varietat d'arquitectures. El pas de codificació de text es pot realitzar amb una xarxa neuronal recurrent, com ara una xarxa de memòria a curt termini (LSTM), tot i que els models de transformadors s'han convertit des d'aleshores en una opció més popular. Per al pas de generació d'imatges, s'han utilitzat habitualment xarxes adversàries generatives condicionals (GAN), i els models de difusió també s'han convertit en una opció popular en els últims anys. En lloc d'entrenar directament un model per

generar una imatge d'alta resolució condicionada a una incrustació de text, una tècnica popular és entrenar un model per generar imatges de baixa resolució i utilitzar un o més models auxiliars d'aprenentatge profund per augmentar-lo, omplint-lo amb detalls de més precisió.

Els models de text a imatge s'entrenen en grans conjunts de dades de parells (text, imatge), sovint obtinguts del web. Amb el seu model **Imagen** de 2022, Google Brain va informar resultats positius de l'ús d'un gran model de llenguatge entrenat per separat en un corpus només de text (amb els seus pesos posteriorment congelats), una desviació de l'enfocament estàndard fins aleshores.

Conjunts de dades

L'entrenament d'un model de text a imatge requereix un conjunt de dades d'imatges combinades amb subtítols de text. Un conjunt de dades que s'utilitza habitualment per a aquest propòsit és el conjunt de dades COCO. Publicat per Microsoft el 2014, COCO consta d'unes 123.000 imatges que representen una diversitat d'objectes amb cinc subtítols per imatge, generades per anotadors humans. Oxford-120 Flowers i CUB-200 Birds són conjunts de dades més petits d'unes 10.000 imatges cadascun, restringits a flors i ocells, respectivament. Es considera menys difícil entrenar un model de text a imatge d'alta qualitat amb aquests conjunts de dades a causa del seu ventall de temes restringit.

Avaluació de la qualitat

Avaluar i comparar la qualitat dels models de text a imatge és un problema que implica avaluar múltiples propietats desitjables. Un criteri específic dels models de text a imatge és que les imatges generades s'alineen semànticament amb els subtítols de text utilitzats per generar-les. S'han dissenyat una sèrie d'esquemes per avaluar aquestes qualitats, alguns d'automatitzats i uns altres de basats en el judici humà.

Una mètrica algorítmica comuna per avaluar la qualitat i la diversitat de la imatge és l'Inception Score (IS), que es basa en la distribució d'etiquetes predites per un model de classificació d'imatges Inceptionv3 prèviament entrenat quan s'aplica a una mostra d'imatges generades pel model text a imatge. La puntuació augmenta quan el model de classificació d'imatges prediu una sola etiqueta amb alta probabilitat, un esquema destinat a afavorir les imatges generades "distintes". Una altra mètrica popular és la distància de Fréchet, que compara la distribució de les imatges generades i les imatges d'entrenament reals segons les característiques extrems per una de les capes finals d'un model de classificació d'imatges prèviament entrenat.

Impacte i aplicacions

La IA té potencial per a la transformació social, cosa que pot incloure l'expansió de gèneres nínjol no comercials per part d'amateurs, noves formes d'entreteniment, prototipatge ràpida, incrementant l'accessibilitat de la producció artística, amb uns resultats més grans per esforç, despeses o temps. Les imatges generades de vegades s'usen com a esbossos, experiments de baix cost, font d'inspiració o idees en fase de prova de concepte. Hi pot haver millores posteriors mitjançant una edició manual posterior, per exemple amb un editor d'imatge.

Llista de models de text a imatge notables

Nom	Data de publicació	Desenvolupador	Llicència
DALL-E	January 2021		
DALL-E 2	April 2022	OpenAI	
DALL-E 3	September 2023		
Ideogram 2.0	August 2024	Ideogram	
Imagen			Propietària
Imagen 2	December 2023 ^[22]	Google	
Parti	Unreleased		
Firefly	June 2023	Adobe Inc.	
Midjourney	July 2022	Midjourney, Inc.	
Stable Diffusion	August 2022	Stability AI	CreativeML Open RAIL-M
RunwayML	2018	Runway AI, Inc.	Propietària

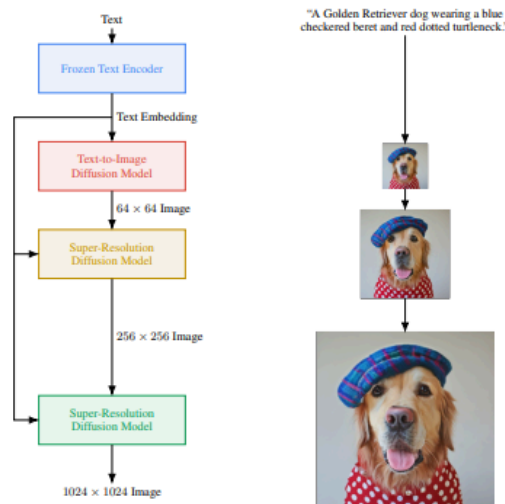


Figure A.4: Visualization of Imagen. Imagen uses a frozen text encoder to encode the input text into text embeddings. A conditional diffusion model maps the text embedding into a 64×64 image. Imagen further utilizes text-conditional super-resolution diffusion models to upsample the image, first $64 \times 64 \rightarrow 256 \times 256$, and then $256 \times 256 \rightarrow 1024 \times 1024$.

https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aeadae0b7d230fa35cbaf04c041-Supplemental-Conference.pdf



Imagen 2 generated this image with the prompt: "Close-up shot of an elderly woman holding a dog in her arms walking on the streets of Prague; capture the fine details of the woman's weathered face, her gentle smile, and the dog's contented expression."

<https://blog.google/technology/ai/google-imagen-2/>

3.4. Àudio i música

La IA generativa també es pot entrenar en clips d'àudio per produir síntesi de veu de so natural i capacitats de text a veu, exemplificades per les eines de síntesi conscients del context d'ElevenLabs o la caixa de veu de Meta Platform.

Els sistemes d'IA generativa com MusicLM i MusicGen també es poden entrenar sobre les formes d'ona d'àudio de la música gravada juntament amb anotacions de text, per tal de generar noves mostres musicals basades en descripcions de text.

Ja s'han generat deepfakes d'àudio de lletres. Els instruments i les lletres de l'artista musical tenen drets d'autor, però les seves veus encara no estan protegides de la IA regenerativa, la qual cosa suscita un debat sobre si els artistes haurien de rebre drets d'autor dels deepfakes d'àudio.

S'han creat molts generadors de música d'IA que es poden generar mitjançant una frase de text, opcions de gènere i biblioteques en bucle de compasos i riffs.

3.5. Text a vídeo

La IA generativa entrenada amb vídeos anotats pot generar clips de vídeo coherents temporalment, detallats i fotorealistes.

Els sistemes més avançats són els següents.

- Sora d'OpenAI <https://openai.com/index/sora/>
- Gen-3 de Runway <https://runwayml.com/research/introducing-gen-3-alpha>
- Veo 2 de Google <https://deepmind.google/technologies/veo/veo-2/>
- Movie Gen de Meta <https://ai.meta.com/research/movie-gen/>
- Kling https://www.youtube.com/@Kling_ai

3.6. Accions

La IA generativa també es pot entrenar sobre els moviments d'un sistema robòtic per generar noves trajectòries per a la planificació del moviment o la navegació. Per exemple, **UniPi** de **Google Research** utilitza indicacions com "recull un bol blau" o "neteja el plat amb una esponja groga" per controlar els moviments d'un braç robot. Els models multimodals de "visió-llenguatge-acció" com ara l'RT-2 de Google poden realitzar raonaments rudimentaris en resposta a les indicacions de l'usuari i a les entrades visuals, com ara recollir un dinosaure de joguina quan s'hi dona la indicació "recull l'animal extingit" en una taula plena de joguines, animals i altres objectes.

3.7. Modelització 3D

El disseny assistit per ordinador (CAD) artificialment intelligent pot utilitzar text-a-3D, imatge-a-3D i vídeo-a-3D per automatitzar el modelatge 3D. Les biblioteques CAD basades en IA també es poden desenvolupar mitjançant dades obertes enllaçades d'esquemes i diagrames. Els assistents CAD AI s'utilitzen com a eines per ajudar a racionalitzar el flux de treball.

4. Programari i maquinari

Els models d'IA generativa s'utilitzen per alimentar diversos agents, inclosos productes de chatbot com ara ChatGPT, eines de programació com GitHub Copilot, productes de text a imatge com Midjourney i productes de text a vídeo com Runway Gen-2. Les funcions d'IA generativa s'han integrat en una varietat de productes comercials existents com Microsoft Office, Google Photos i Adobe Photoshop. Molts models d'IA generativa també estan disponibles com a programari de codi obert, com Stable Diffusion i el model de llenguatge LLaMA.

Els models d'IA generativa més petits amb uns quants milers de milions de paràmetres es poden executar en telèfons intel·ligents, dispositius incrustats i ordinadors personals. Per exemple, LLaMA-7B (una versió amb 7.000 milions de paràmetres) es pot executar en un Raspberry Pi 4 i una versió de Stable Diffusion es pot executar en un iPhone 11.

Els models més grans amb desenes de milers de milions de paràmetres poden funcionar en ordinadors portàtils o d'escriptori. Per aconseguir una velocitat acceptable, els models d'aquesta mida poden requerir acceleradors com els xips GPU produïts per NVIDIA i AMD o el Neural Engine inclòs als productes de silici d'Apple. Per exemple, la versió de 65.000 milions de paràmetres de LLaMA es pot configurar per executar-se en un ordinador d'escriptori.

Els avantatges d'executar IA generativa localment inclouen la protecció de la privadesa i la propietat intel·lectual, i evitar la limitació de tarifes i la censura. El subreddit r/LocalLLaMA, en particular, se centra en l'ús de targetes gràfiques de jocs de qualitat mitjançant tècniques com la compressió. Aquest fòrum és una de les dues úniques fonts en què confia Andrej Karpathy per a referents de models lingüístics. Yann LeCun ha defensat els models de codi obert pel seu valor per a aplicacions verticals i per millorar la seguretat de la IA.

Els models d'idioma amb centenars de milers de milions de paràmetres, com ara GPT-4 o PaLM, s'executen normalment en ordinadors de centres de dades equipats amb matrius de GPU (com l'H100 de NVIDIA) o xips acceleradors d'IA (com ara el TPU de Google). Normalment s'accedeix a aquests models molt grans com a serveis al núvol a través d'Internet.

El 2022, els nous controls d'exportació dels Estats Units en informàtica avançada i semiconductors a la Xina van imposar restriccions a les exportacions a la Xina de xips d'acceleració de GPU i IA utilitzats per a IA generativa. Xips com el NVIDIA A800 i el Biren Technology BR104 es van desenvolupar per complir els requisits de les sancions.

Hi ha programari lliure al mercat capaç de reconèixer el text generat per la intel·ligència artificial generativa (com GPTZero), així com les imatges, l'àudio o el vídeo que en provenen. Malgrat les afirmacions de precisió, els detectors de text d'IA gratuïts i de pagament sovint han produït falsos positius, acusant erròniament els estudiants d'enviar treballs generats per IA.

5. Regulació

Als Estats Units, un grup d'empreses com OpenAI, Alphabet i Meta va signar un acord voluntari amb la Casa Blanca el juliol de 2023 per marcar el contingut generat per IA. L'octubre de 2023, l'Ordre executiva 14110 va aplicar la Llei de producció de defensa per exigir a totes les empreses nord-americanes que informin d'informació al govern federal quan entrenen grans models d'IA.

A la Unió Europea, la Llei d'Intel·ligència Artificial proposada inclou requisits per revelar material amb drets d'autor utilitzat per entrenar sistemes d'IA generativa i etiquetar com a tal qualsevol sortida generada per IA.

La regulació de la intel·ligència artificial ha adquirit més importància en les últimes setmanes. L'administració de Biden va presentar dilluns una nova ordre executiva integral amb la intenció de canviar la manera com el govern federal aborda la intel·ligència artificial. Aquesta directiva, entre altres coses, requereix que les empreses creïn models específics d'IA generativa d'alt impacte per informar el govern i revelar les conclusions de les seves proves. A mesura que la UE treballa per convertir-se en el principal regulador mundial de la intel·ligència artificial, el Regne Unit inicia la seva "cimera de seguretat de la intel·ligència artificial". Limitar els riscos associats a la IA generativa s'està convertint en una preocupació més urgent d'aquestes noves solucions.

A la Xina, les mesures provisionals per a la gestió dels serveis d'IA generativa introduïdes per l'Administració del ciberespai de la Xina regulen qualsevol intel·ligència artificial generativa de cara al públic. Inclou requisits sobre imatges o vídeos generats, regulacions sobre dades de formació i qualitat de les etiquetes, restriccions a la recollida de dades personals i una directriu que la IA generativa ha de "adherir-se als valors fonamentals socialistes".

Entrenament amb contingut protegit per drets d'autor

Els sistemes d'IA generativa, com ara ChatGPT i Midjourney, s'entrenen en grans conjunts de dades disponibles públicament que inclouen obres amb drets d'autor. Els desenvolupadors d'IA han argumentat que aquesta formació està protegida per un ús legítim, mentre que els titulars dels drets d'autor han argumentat que infringeix els seus drets.

Els defensors de la formació d'ús just han argumentat que és un ús transformador i no implica posar còpies d'obres amb drets d'autor a disposició del públic. Els crítics han argumentat que els generadors d'imatges com Midjourney poden crear còpies gairebé idèntiques d'algunes imatges amb drets d'autor, i que els programes d'IA generativa competeixen amb el contingut en què s'entrenen.

A partir del 2024, hi ha diverses demandes relacionades amb l'ús de material amb drets d'autor en entrenament. Getty Images ha demandat Stability AI per l'ús de les seves imatges per entrenar la difusió estable. Tant l'Authors Guild com el New York Times han demandat Microsoft i OpenAI per l'ús de les seves obres per entrenar ChatGPT.

Copyright del contingut generat per IA

Una altra pregunta és si les obres generades per IA poden optar a la protecció dels drets d'autor. L'Oficina de drets d'autor dels Estats Units ha dictaminat que les obres creades per intel·ligència artificial sense cap intervenció humana no poden tenir drets d'autor, perquè no tenen autoria humana. Tanmateix, l'oficina també ha començat a rebre aportacions públiques per determinar si aquestes regles s'han de perfeccionar per a la IA generativa.

6. Qüestions ètiques i legals

El desenvolupament de la IA generativa ha suscitat preocupacions per part dels governs, les empreses i els individus, cosa que ha provocat protestes, accions legals, crides a posar en pausa els experiments d'IA i accions de diversos governs. En una sessió informativa de juliol de 2023 del Consell de Seguretat de les Nacions Unides, el secretari general António Guterres va declarar que "la IA generadora té un enorme potencial per al bé i el mal a escala", que la IA pot "turbocarregar el desenvolupament global" i contribuir entre 10 i 15 bilions de dòlars al món global. economia per al 2030, però que el seu ús maliciós "podria causar nivells horribles de mort i destrucció, trauma generalitzat i danys psicològics profunds a una escala inimaginable".

Pèrdues de llocs de treball

Tot i que no era una prioritat màxima, una de les sol·licituds de la WGA de 2023 va ser "regulacions sobre l'ús de la IA (generativa)".

Des dels primers dies del desenvolupament de la IA, el creador d'ELIZA, Joseph Weizenbaum i altres, han presentat arguments sobre si les tasques que poden fer els ordinadors realment les haurien de fer ells, donada la diferència entre ordinadors i humans, i entre quantitats. càlculs i judicis qualitius basats en valors. L'abril de 2023, es va informar que la IA de generació d'imatges va provocar la pèrdua del 70% dels llocs de treball dels il·lustradors de videojocs a la Xina. El juliol de 2023, l'evolució de la IA generativa va contribuir als conflictes laborals de Hollywood de 2023. Fran Drescher, president del Screen Actors Guild, va declarar que "la intel·ligència artificial suposa una amenaça existencial per a les professions creatives" durant la vaga SAG-AFTRA de 2023. La IA de generació de veu s'ha vist com un repte potencial per al sector de la veu.

La intersecció de la IA i les preocupacions laborals entre grups subrepresentats a nivell mundial continua sent una faceta crítica. Tot i que la IA promet millores de l'eficiència i l'adquisició d'habilitats, les preocupacions sobre el desplaçament de llocs de treball i els processos de contractació esbiaixats persisteixen entre aquests grups, tal com es descriuen a les enquestes de Fast Company. Per aprofitar la IA per a una societat més equitativa, els passos proactius inclouen mitigar els biaixos, defensar la transparència, respectar la privadesa i el consentiment i adoptar diversos equips i consideracions ètiques. Les estratègies impliquen reorientar l'èmfasi de les polítiques en la regulació, el disseny inclusiu i el potencial de l'educació per a l'ensenyament personalitzat per maximitzar els beneficis alhora que es minimitzen els danys.

Biaix racial i de gènere

Els models d'IA generativa poden reflectir i amplificar qualsevol biaix cultural present a les dades subjacents. Per exemple, un model lingüístic podria suposar que els metges i els jutges són homes i que les secretàries o les infermeres són dones, si aquests biaixos són habituals a les dades de formació. De la mateixa manera, un model d'imatge sol·licitat amb el text "una foto d'un CEO" podria generar de manera desproporcionada imatges de consellers delegats masculins blancs, si s'entrena en un conjunt de dades esbiaixat racialment. S'han intentat diversos mètodes per mitigar el biaix, com ara modificar les indicacions d'entrada i tornar a ponderar les dades d'entrenament.

Deepfakes

Els deepfakes són mitjans generats per IA que prenen una persona en una imatge o un vídeo existent i els substitueixen per la imatge d'una altra persona mitjançant xarxes neuronals artificials. Els deepfakes han despertat una gran atenció i preocupació pel seu ús en vídeos pornogràfics de celebritats, pornografia de venjança, notícies falses, enganys, desinformació sanitària i frau financer. Això ha provocat respostes tant de la indústria com del govern per detectar i limitar-ne l'ús.

Deepfakes d'àudio

Les instàncies d'usuaris que abusen del programari per generar declaracions controvertides en l'estil vocal de celebritats, funcionaris públics i altres persones famoses han plantejat preocupacions ètiques sobre la IA de generació de veu. En resposta, empreses com ElevenLabs han afirmat que treballarien per mitigar els possibles abusos mitjançant salvaguardes i verificació d'identitat.

Les preocupacions i el fandom han sorgit de la música generada per IA. El mateix programari que s'utilitza per clonar veus s'ha utilitzat en veus de músics famosos per crear cançons que imiten les seves veus, guanyant una enorme popularitat i crítiques. També s'han utilitzat tècniques similars per crear versions de llarga durada o de qualitat millorada de cançons que s'han filtrat o encara no s'han publicat.

La intel·ligència artificial generativa també s'ha utilitzat per crear noves personalitats d'artistes digitals, algunes d'elles reben prou atenció com per rebre acords discogràfics a les grans discogràfiques. Els desenvolupadors d'aquests artistes virtuals també s'han enfrontat a la seva part justa de crítiques pels seus programes personificats, inclosa la reacció per "deshumanitzar" una forma d'art i també per crear artistes que creen atractius irreals o immorals per al seu públic.

La ciberdelinqüència

La capacitat de l'IA generativa per crear contingut fals realista s'ha explotat en nombrosos tipus de ciberdelictes, incloses les estafes de pesca. El vídeo i l'àudio deepfake s'han utilitzat per crear desinformació i frau. L'antic tsar del frau de Google, Shuman Ghosemajumder, ha pronosticat que, si bé els vídeos deepfake inicialment van crear un enrenou als mitjans, aviat es convertirien en un lloc habitual i, com a resultat, més perillosos. A més, s'han utilitzat models en llenguatge gran i altres formes d'IA de generació de text a gran escala per crear ressenyes falses en llocs web de comerç electrònic per augmentar les valoracions. Els ciberdelinqüents han creat grans models de llenguatge centrats en el frau, com WormGPT i FraudGPT.

Una investigació recent feta el 2023 ha revelat que la IA generativa té debilitats que els delinqüents poden manipular per extreure informació nociva sense passar les garanties ètiques. L'estudi presenta exemples d'atacs realitzats a ChatGPT, inclosos jailbreaks i psicologia inversa. A més, els individus maliciosos poden utilitzar ChatGPT per a atacs d'enginyeria social i atacs de pesca, revelant el costat nociu d'aquestes tecnologies.

Mal ús en el periodisme

El gener de 2023, Futurism.com va publicar la història que CNET havia estat utilitzant una eina d'IA interna no revelada per escriure almenys 77 de les seves històries; després de la notícia, CNET va publicar correccions a 41 de les històries.

L'abril de 2023, el tabloide alemany Die Aktuelle va publicar una falsa entrevista generada per IA amb l'antic pilot de carreres Michael Schumacher, que no havia fet cap aparició pública des del 2013 després de patir una lesió cerebral en un accident d'esquí. La història incloïa dues possibles revelacions: la portada incloïa la línia "enganyosament real", i l'entrevista incloïa un reconeixement al final que era generada per IA. L'editor en cap va ser acomiadat poc després enmig de la polèmica.

Altres publicacions que han publicat articles el contingut i/o l'autorització dels quals s'ha confirmat o sospita que ha estat creat per models d'IA generativa, sovint amb contingut fals, errors i/o no divulgació de l'ús generatiu d'IA, inclouen NewsBreak, publicacions propietat d'Arena Group (Sports Illustrated, TheStreet, Men's Journal), B&H Photo, publicació propietat de Gannett (The Columbus Dispatch revisat), MSN, News Corp, publicacions propietat de G/O Media (Gizmodo, Jalopnik, A.V. Club), The Irish Times, publicacions propietat de Red Ventures (Bankrate) i BuzzFeed.

En resposta als possibles inconvenients relacionats amb l'ús i el mal ús de la IA generativa en el periodisme, mitjans com Wired, The Associated Press i The Guardian han publicat directrius sobre com planegen utilitzar i no utilitzar la IA generativa en el seu treball.

7. Tothom pot ser artista?

Manllevam el títol d'aquest apartat del capítol 58 del llibre de López de Mántaras. L'autor hi fa una comparació interessant entre IA i fotografia referida a la relació entre tecnologia i art. Hi explica la següent història.

Al setembre de 2022, Jason Allen, un dissenyador de jocs, va utilitzar Midjourney, un programari d'IA que converteix descripcions de text en imatges, per crear *Théâtre d'Opéra Spatial*. És una obra estèticament bella que mostra un encertat equilibri entre l'estil d'òpera barroca tradicional, reflectida a la roba de la gent, i les característiques arquitectòniques de ciència-ficció del film *Dune*. És un exemple més d'art generat usant IA amb la particularitat que ha guanyat un primer premi en un concurs de belles arts a Colorado en la categoria d'art digital per a artistes emergents. No és una gran proesa, però potser ens dona una pista del que vindrà.

Això va tenir repercussió pública.

Diversos mitjans de comunicació importants van cobrir la notícia i això ha generat enceses discussions a les xarxes. Els arguments més comuns que la gent ha utilitzat contra l'art generat mitjançant IA, i els artistes que utilitzen aquestes eines, són: "qualsevol pot utilitzar la IA per crear imatges, però això no els converteix en artistes", "no cal cap habilitat per utilitzar aquestes eines" i la crítica més comuna: "l'art de la IA no és art".

Com varen ser les primeres proves?

Diverses persones, inclosos artistes, que van provar Midjourney es van adonar que podien obtenir fàcilment obres més aviat mediocres amb descripcions textuais senzilles, delegant efectivament la major part del treball a la IA. Per tant, no tindria sentit dir que estaven "creant" una obra d'art. Ara bé, la majoria va admetre que no sabrien com generar obres realment belles estèticament sense dedicar-hi un bon grapat d'hores. Efectivament, cal desenvolupar habilitats importants per aconseguir grans resultats. El mateix Allen va generar molts centenars d'obres que va analitzar detingudament i en va seleccionar només tres. Finalment, va millorar-ne significativament una, afegint-hi fins i tot elements importants, com ara el cap d'algun dels personatges i canviant les proporcions i colors de diversos elements de l'obra. És cert que és més fàcil adquirir l'habilitat necessària per obtenir bons resultats amb Midjourney que adquirir l'habilitat per crear una pintura a l'oli sobre tela que guanyi un concurs. No obstant això, aquesta nova forma d'art amb IA també demana una habilitat no gens trivial. Per tant, no és cert que l'art amb IA no requereix cap habilitat i, per tant, no és cert que tothom pugui crear art amb IA.

Podem comparar-ho amb l'adopció artística d'una tècnica anterior?

La situació actual no és tan diferent de la que es va donar amb una altra tecnologia que va aparèixer fa més de cent vuitanta anys: la fotografia. Les eines d'art amb IA comparteixen amb la fotografia més que la seva capacitat d'influir en la nostra comprensió de l'art. La fotografia no es va inventar amb la finalitat de fer art. Era una manera mecànica d'enregistrar moments fugaçs de la vida, fent realitat el que Goethe va fer desitjar al seu Faust: "aturar un instant i preservar-lo per sempre". Era una qüestió d'habilitat tècnica. Aquí és on comencen els paral·lelismes amb els models d'art amb IA: el conjunt d'habilitats necessàries per fer funcionar les primeres càmeres no eren artístiques, sinó tècniques. Va ser només amb el pas del temps que ambdós tipus d'habilitats es van fusionar. De fet, avui en dia saber fer impressionants fotografies requereix més habilitat artística que habilitat tècnica.

Els primers fotògrafs van haver d'esbrinar els efectes en el resultat final d'aspectes com ara la posició, la il·luminació, l'angle, la profunditat de camp o la distància focal. Van explorar els territoris inexplorats de la fotografia i finalment van descobrir les possibilitats pràcticament infinites que ofereix aquest art.

L'autor acaba amb la predicció següent.

Passarà el mateix amb l'art amb IA. Només aquells que estiguin disposats a fer l'esforç d'entendre les tècniques d'IA, a explorar el que és possible crear amb elles i a adquirir les habilitats necessàries per fer-ho són els que finalment anomenarem artistes. A mesura que aquesta tecnologia evolucioni amb el temps, l'art amb IA prendrà lentament la forma d'una nova categoria d'art, com va passar amb la fotografia fa més de cent vuitanta anys.

Hi estau d'acord?

8. Enginyeria de prompt

Vegem a continuació sis principis generals d'enginyeria de prompting que ens seran útils per treure més profit de les eines d'IA generativa.

<https://cloud.google.com/blog/products/application-development/five-best-practices-for-prompt-engineering>

8.1. Conèixer el model

A mesura que els models d'IA evolucionen i es tornen més complexos, és essencial que els desenvolupadors entenguin les seves capacitats i limitacions. Comprendre aquests punts forts i febles us pot ajudar, com a desenvolupador, a evitar cometre errors i crear aplicacions més segures i fiables.

Per exemple, un model d'IA que està entrenat per reconèixer imatges de nabis pot no ser capaç de reconèixer imatges de maduixes. Per què? Perquè el model només es va entrenar en un conjunt de dades d'imatges de nabis. Si un desenvolupador utilitza aquest model per crear una aplicació que se suposa que reconeix tant els nabis com les maduixes, és probable que l'aplicació cometi errors, la qual cosa condueix a un resultat ineficaç i una mala experiència d'usuari.

És important tenir en compte que els models d'IA tenen la capacitat de ser esbiaixats. Això es deu al fet que s'entrenen models d'IA amb dades que es recullen del món real i, per tant, poden reflectir la dinàmica de poder desigual arrelada inherentment a la nostra jerarquia social. Si les dades que s'utilitzen per entrenar un model d'IA estan esbiaixades, el model també estarà esbiaixat. Això pot provocar problemes si el model s'utilitza per prendre decisions que afecten les persones reforçant els biaixos socials. Abordar aquests biaixos és important per garantir que les dades siguin justes, promovent la igualtat i garantint la responsabilitat de la tecnologia d'IA. Els enginyers prompts haurien de ser conscients de les limitacions o biaixos de la formació perquè puguin elaborar indicacions de manera més eficaç i entendre quin tipus d'indicacions és possible fins i tot per a un model determinat.

8.2. Especificitat

Els models d'IA tenen la capacitat de comprendre una varietat d'indicacions. Tot i que els models d'IA poden tenir molt coneixements, encara són imperfectes i tenen la capacitat d'interpretar malament les indicacions que no són prou específiques. Perquè els models d'IA puguin navegar per l'ambigüitat, és important adaptar les indicacions específicament al resultat desitjat.

Suposem que us agradaria que el vostre model d'IA generi una recepta per a 50 magdalenes de nabius vegans. Si demaneu al model "quina és una recepta de magdalenes de nabius?", el model no sap que necessiteu fer 50 magdalenes. Per tant, és poc probable que enumereu el volum més gran d'ingredients que necessitareu o inclogueu consells per ajudar-vos a coure de manera més eficient un nombre tan gran de magdalenes. El model només pot sortir del context proporcionat. Una indicació més eficaç seria "Estic allotjant 50 convidats. Genereu una recepta per a 50 magdalenes de nabius". És més probable que el model generi una resposta que sigui rellevant per a la vostra sol·licitud i que compleixi els vostres requisits específics.

8.3. Prompts contextuais

Utilitzeu informació contextual a les vostres sol·licituds per ajudar el model a obtenir una comprensió profunda de les vostres sol·licituds. Les indicacions contextuais poden incloure la tasca específica que voleu que realitzi el model, una rèplica de la sortida que esteu buscant o una persona per emular, des d'un venedor o enginyer fins a un professor de secundària. La definició d'un to i una perspectiva per a un model d'IA li dona un esquema del to, l'estil i l'experiència enfocada que busqueu per millorar la qualitat, la rellevància i l'eficàcia de la vostra sortida.

En el cas de les magdalenes de nabius, és important indicar el model utilitzant el context de la situació. El model podria necessitar més context que generar una recepta per a 50 persones. Si ha de ser conscient que la recepta ha de ser vegana, podeu demanar-li al model que respongui imitant un xef vegà hàbil.

En proporcionar indicacions contextuais, podeu ajudar a garantir que les vostres interaccions amb IA siguin tan fluides i eficients com sigui possible. El model podrà entendre més ràpidament la teva sol·licitud i podrà generar respostes més precises i rellevants.

8.4. Exemples

Quan es creen sol·licituds per a models d'IA, és útil proporcionar exemples. Això es deu al fet que les indicacions actuen com a instruccions per al model i els exemples poden ajudar el model a entendre el que esteu demanant. Proporcionar una indicació amb un exemple sembla una cosa així: "aquí hi ha diverses receptes que m'agraden: creeu una recepta nova basada en les que he proporcionat". El model ara pot entendre les vostres habilitats i necessitats per fer aquesta pastisseria,

8.5. Prompts i personatges

La forma en què construïu la vostra indicació afecta la sortida del model. En explorar de manera creativa diferents sol·licituds, aviat entendreu com el model pondera les seves respostes i què passa quan fusioneu el vostre coneixement del domini, l'experiència i l'experiència viscuda amb la potència d'un model d'idioma gran amb milers de milions de paràmetres.

Proveu d'experimentar amb diferents paraules clau, estructures de frases i longituds d'indicacions per descobrir la fórmula perfecta. Permeteu-vos posar-vos a la pell de diverses persones, des de persones de feina com ara "enginyer de producte" o "representants d'atenció al client", fins a figures parentals o celebritats com la vostra àvia, un xef famós, i exploreu tot, des de la cuina fins a la codificació!

En crear sol·licituds úniques i innovadores plenes de la vostra experiència i experiència, podeu saber quines indicacions us proporcionen el vostre resultat ideal. Perfeccionar encara més les vostres indicacions, conegudes com a "ajustament", permet que el model tingui una major comprensió i marc per a la vostra propera sortida.

8.6. Chain of Thought

La cadena de pensament (en anglès *Chain of Thought*) és una tècnica per millorar les capacitats de raonament dels grans models de llenguatge (LLM). Funciona desglossant un problema complex en passos més petits i, a continuació, demana al LLM que proporcioni un raonament intermedi per a cada pas. Això ajuda al LLM a comprendre el problema més profundament i a generar respostes més precises i informatives. Això us ajudarà a entendre millor la resposta i a assegurar-vos que el LLM realment entén el problema.

8.7. Conclusió

L'enginyeria de prompt és una habilitat que necessitaran tots els treballadors cada vegada més, en totes les indústries i organitzacions, ja que les eines impulsades per IA són cada cop més freqüents.

Podeu incorporar aquestes directrius essencials la propera vegada que interactueu amb un model d'IA, perquè pugueu generar les sortides precises que desitjau.

9. Casos d'ús

En aquest apartat presentarem diversos casos d'ús d'IA generativa, sovint combinant diferents modalitats.

9.1. Ganyet - Mazoni



Entre el CEO de Mortensen, Josep Maria Ganyet, i el músic i productor Jaume Pla **Mazoni** van mostrar com es pot crear un tema musical incloent ritme, melodia, harmonia, instrumentació i estructura i lletra del tema amb l'ajuda de la IA.

<https://www.viasona.cat/noticia/4462/arriba-xuixos-cosmics-una-canço-generada-amb-ajuda-de-la-ia>

9.2. Boris Eldagsen



Boris Eldagsen, un fotògraf i artista alemany, va guanyar la categoria creativa oberta als Premis Sony World Photography Awards del 2023, però va rebutjar el premi després de revelar que la seva imatge guanyadora havia estat generada per intel·ligència artificial. L'obra d'Eldagsen, titulada "Pseudomnesia: The Electrician", mostrava un inquietant retrat en blanc i negre de dues dones de diferents generacions.

- Eldagsen va presentar la imatge generada per IA al concurs com una prova deliberada per veure si aquests esdeveniments estaven preparats per gestionar contingut generat per IA.
- Es va descriure a si mateix com un "mico entremaliat" per participar al concurs amb una imatge d'IA, amb l'objectiu de provocar un debat sobre el paper de la IA en la fotografia.
- Després de guanyar el premi, Eldagsen el va rebutjar, afirmant que "les imatges d'IA i la fotografia no haurien de competir entre si en un premi com aquest. Són entitats diferents. La IA no és fotografia."

Les accions d'Eldagsen han encès un debat sobre el futur de la fotografia i la creixent dificultat per diferenciar entre fotografies originals i imatges generades per IA. Aquest incident planteja qüestions importants sobre l'ètica, l'autenticitat i el valor artístic en l'era de la intel·ligència artificial.

<https://www.bbc.com/news/entertainment-arts-65296763>

9.3. DreamUp

DreamUp és una eina de generació d'imatges amb intel·ligència artificial creada per DeviantArt. Permet als usuaris crear art, oferint 5 indicacions gratuïtes per començar. Les imatges generades s'etiqueten automàticament com a #Alart i es dona crèdit als creadors si el seu estil s'utilitza com a inspiració.

Les característiques principals de DreamUp són les següents:

- Genera imatges d'alta resolució basades en descripcions de text
- Permet als usuaris controlar com s'utilitza el seu art i si volen que s'entreni models d'IA amb el seu contingut
- Integra les imatges generades a la plataforma DeviantArt
- Ofereix opcions per millorar la resolució i explorar variacions de les imatges creades

<https://www.deviantart.com/dreamup>

10. Un univers en expansió

L'aparició d'eines d'IA generativa és explosiva. Una gran quantitat de noves marques apareix contínuament i és ben difícil seguir el flux d'informació nova i experimentar amb un ventall ampli d'elles.

Aquesta és una mostra d'eines al moment d'elaborar els apunts, construïda amb l'ajut de Perplexity.

<https://www.perplexity.ai/search/i-have-the-following-use-cases-1UwW0zsMSTaXwtWQp80NmA>

11. Reptes

L'ús de la IA generativa està generant moltes notícies, que reflecteixen els problemes d'aquesta tecnologia pel que fa a la privacitat, el dret a la pròpia imatge, o l'impacte laboral, per exemple.

A continuació n'ofereixo una mostra.

- Deepfake contra la intimitat <https://www.vilaweb.cat/noticies/detingut-palma-menor-imatges-companyes-classe-nues-intelligencia-artificial/>
- Autoria, creació artística <https://www.vilaweb.cat/noticies/brutalista-ia-hollywood-controversia-oscars/>
- Impacte de la IA generativa a l'activitat laboral <https://www.vilaweb.cat/noticies/dolorosa-veritat-rethink-x-feina-tots/>
- L'Estat multarà qui no etiqueti contingut d'IA <https://www.elpuntavui.cat/societat/article/15-ciencia/2516832-l-estat-multara-qui-no-etiqueti-contingut-d-ia.html>

12. Enllaços

Podeu aprofundir el coneixement de la IA generativa consultant els enllaços següents.

- <https://aws.amazon.com/generative-ai/>

AWS ofereix la seva solució completa per a la implementació d'IA generativa, proporcionant eines per construir i escalar aplicacions utilitzant models de llenguatge extensos (LLM) i models fonamentals (FM). La plataforma dona accés a models líders en la indústria, infraestructura optimitzada i aplicacions potenciades per IA generativa, abordant diversos casos d'ús com xatbots, anàlisi conversacional, assistents per a empleats, generació de codi i personalització. A més de les seves capacitats tècniques, AWS emfatitza el seu compromís amb el desenvolupament responsable de la IA i ofereix recursos educatius com PartyRock i una comunitat de desenvolupadors. La plataforma també introdueix Amazon Q, una eina dissenyada per millorar la productivitat empresarial i l'experiència del client. En conjunt, AWS presenta una oferta integral per a organitzacions que busquen innovar en el camp de la IA generativa, oferint les eines, la infraestructura i els recursos necessaris per aprofitar aquesta tecnologia transformadora.

- <https://keras.io/examples/generative/>

Keras ofereix una àmplia col·lecció d'exemples i tutorials centrats en l'aprenentatge profund generatiu, abastant diverses tècniques i aplicacions. Aquests recursos cobreixen una gamma diversa de temes, des de models de difusió com Denoising Diffusion Implicit Models i Stable Diffusion, fins a arquitectures més tradicionals com Variational AutoEncoders i diferents tipus de Generative Adversarial Networks (GANs). També s'inclouen exemples específics per a la generació d'imatges, com Deep Dream, GauGAN i StyleGAN, així com tècniques de transferència d'estil.

A més de la generació d'imatges, Keras proporciona tutorials per a la generació de text utilitzant models com GPT i LSTM, i fins i tot aborda la generació de gràfics moleculars per a aplicacions farmacèutiques. Aquests exemples estan dissenyats per a diferents versions de Keras, incloent-hi la més recent Keras 3, oferint als desenvolupadors una visió completa de com implementar i utilitzar aquestes tècniques d'aprenentatge profund generatiu en els seus propis projectes. La diversitat i profunditat d'aquests recursos fan de Keras una plataforma valuosa per a aquells que busquen explorar i aplicar tècniques d'IA generativa avançades.

- <https://www.tensorflow.org/tutorials/generative>

TensorFlow ofereix una sèrie de tutorials exhaustius sobre models generatius, centrant-se en diverses tècniques avançades d'aprenentatge profund. Aquests tutorials cobreixen una àmplia gamma de temes, incloent-hi la generació d'imatges amb Autoencoders Variacionals (VAE) i Xarxes Generatives Adversarials (GAN), així com la generació de text utilitzant models de llenguatge com RNN i Transformers. Els exemples proporcionen implementacions detallades d'arquitectures populars com DCGAN, Pix2Pix i CycleGAN per a tasques de generació i transformació d'imatges.

A més de les tècniques de generació d'imatges i text, TensorFlow també ofereix tutorials sobre mètodes més avançats com la síntesi de música amb RNN i la generació d'imatges estilitzades. Cada tutorial inclou explicacions pas a pas, codi font complet i visualitzacions dels resultats, permetent als desenvolupadors entendre i implementar aquests models complexos. Aquests recursos són útils per aprofundir en el camp de l'aprenentatge profund generatiu utilitzant el framework de TensorFlow.

- <https://cloud.google.com/ai/generative-ai>

Google Cloud ofereix una plataforma completa per a la implementació de solucions d'IA generativa, centrada en el seu model multimodal Gemini. Aquesta plataforma permet als desenvolupadors i empreses construir aplicacions i experiències basades en IA generativa de manera ràpida, eficient i responsable. A través de Vertex AI, els usuaris poden interactuar amb models fonamentals, personalitzar-los i integrar-los en les seves aplicacions. Gemini, desenvolupat per Google DeepMind, destaca per la seva capacitat de processar diversos tipus d'entrada (text, imatges, vídeo i codi) i generar sortides variades, oferint funcionalitats avançades de raonament i generació.

Google Cloud també proporciona eines específiques per a desenvolupadors, com Gemini Code Assist, que ajuda a construir aplicacions amb major velocitat, qualitat i seguretat en entorns populars com VS Code i JetBrains. A més, la plataforma emfatitza la importància d'un ecosistema obert i innovador de socis d'IA, oferint als clients

opcions i oportunitats d'innovació. Per a les startups, Google ofereix el programa Google for Startups Cloud, que proporciona crèdits substancials i accés a recursos i tecnologies com Vertex AI per accelerar el desenvolupament de projectes d'IA generativa.

- <https://learn.microsoft.com/en-us/training/paths/introduction-generative-ai/>

Microsoft ofereix una ruta d'aprenentatge completa sobre els fonaments de la IA generativa a Azure, dissenyada per proporcionar una comprensió sòlida d'aquesta tecnologia emergent. El curs cobreix els conceptes bàsics dels models de llenguatge que formen la base de la IA generativa, explicant com aquests models poden generar contingut original basat en entrades de llenguatge natural. També s'introdueix Azure AI Studio, una plataforma que proporciona accés a les últimes tecnologies de IA generativa, i s'explora com les aplicacions de IA generativa, com els copilots, poden augmentar l'eficiència en diverses tasques.

Aquesta ruta d'aprenentatge també aborda aspectes pràctics i ètics de la IA generativa. S'expliquen els conceptes de prompts i respostes dels copilots, proporcionant una comprensió de com interactuar eficaçment amb aquestes eines. A més, es fa èmfasi en els principis de IA responsable de Microsoft, destacant la importància de desenvolupar i implementar solucions de IA generativa de manera ètica. Aquest curs està dissenyat per a aquells que ja tenen certa familiaritat amb Azure i el portal d'Azure, oferint una base sòlida per a futurs desenvolupaments en el camp de la IA generativa.