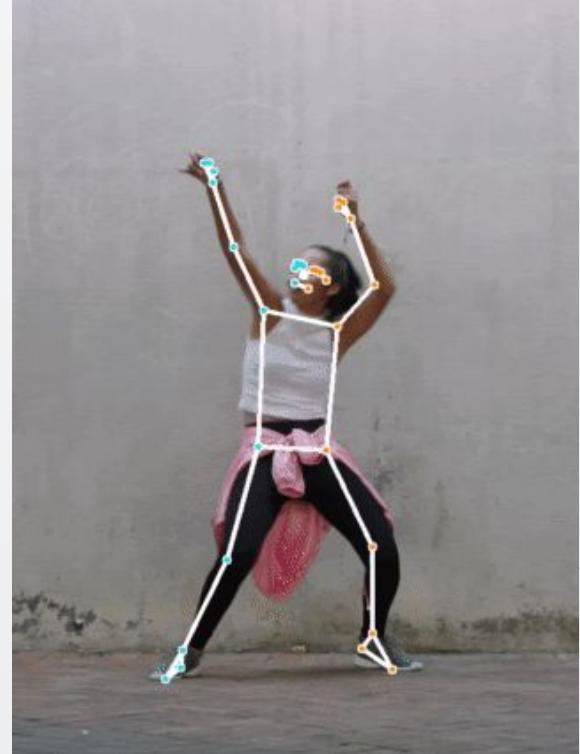


# Desbloqueando el potencial de los datos: Estrategias Avanzadas de Computer Vision y Deep Learning.

MAYO 2024

Cristian Lazo Quispe  
AI technical lead





*Desde que un niño aprende a hablar, comienza a describir el mundo que lo rodea. Hoy, la inteligencia artificial emula esta habilidad, revolucionando la industria con la visión por computadora.*

# TABLA DE CONTENIDO

**01**

## Fundamentos

Descripción de qué es y cómo nace Computer Vision

**02**

## Aplicaciones

Cómo y con qué se trabaja Computer Vision

**03**

## Demo

¡Código en vivo!

**04**

## Ética

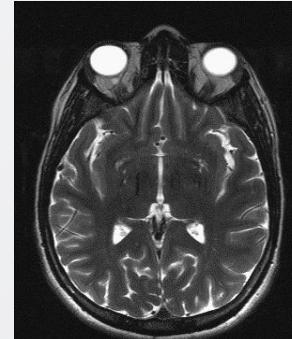
Discusión sobre el futuro

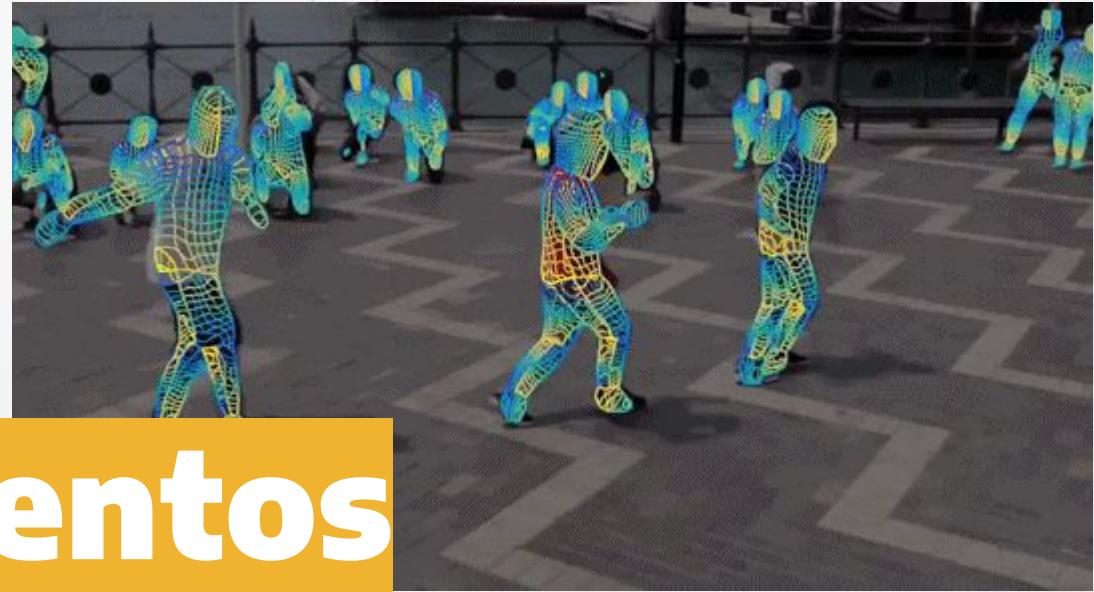


Cristian Lazo Quispe  
AI Technical Lead - Nisum  
ML Research Engineer - UNI, Perú

# Acerca de mi

Soy un innovador autodidacta y un investigador comprometido, siempre buscando nuevas formas de aplicar la IA para generar un impacto positivo en la sociedad.



**01**

# Fundamentos

Descripción de qué es y  
cómo nace Computer Vision

# QUÉ ES COMPUTER VISION?

c

Explicale a un niño de 5 años que es computer vision y para que se usa

¡Claro! Aquí tienes una explicación de Computer Vision para un niño de 5 años:

"¿Sabes cómo tus ojitos te ayudan a ver y reconocer cosas? Puedes ver a mamá, a papá, tus juguetes y todo lo que te rodea. Bueno, imagina que las computadoras también pudieran ver y entender lo que hay en una imagen o un video, casi como si tuvieran ojos mágicos. Eso es lo que llamamos 'Computer Vision' o 'Visión por Computadora'.

# El misterio de la visión

*Psychological Review*  
Vol. 61, No. 3, 1954

## SOME INFORMATIONAL ASPECTS OF VISUAL PERCEPTION

FRED ATTNEAVE

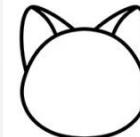
*Perceptual and Motor Skills Research Laboratory,  
Human Resources Research Center<sup>1</sup>*

The ideas of information theory are at present stimulating many different areas of psychological inquiry. In providing techniques for quantifying situations which have hitherto been difficult or impossible to quantify, they suggest new and more precise ways of conceptualizing these situations (see Miller [12] for a general discussion and bibliography). Events ordered in time are particularly amenable to informational analysis; thus language sequences are being extensively studied, and other sequences, such as those of music, plainly

plications, is precisely equivalent to an assertion that the world as we know it is lawful. In the present discussion, however, we shall restrict our attention to special types of lawfulness which may exist in space at a fixed time, and which seem particularly relevant to processes of visual perception.

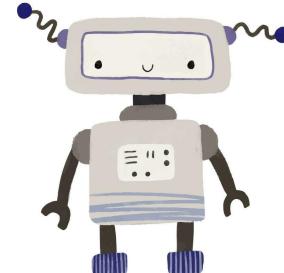
### THE NATURE OF REDUNDANCY IN VISUAL STIMULATION: A DEMONSTRATION

Consider the very simple situation



*La información más compleja con la más simple son equivalentes para nuestra percepción.*

# Entendimiento de lo que vemos



Un grupo de personas está de pie frente a una pizarra, iluminadas por luces en el techo. La sala tiene una pizarra en la pared y un cartel azul con texto blanco. Una mujer con un suéter blanco y un hombre con un suéter azul están entre las personas en la sala. El cartel es azul y el texto es blanco. La mujer con el suéter blanco tiene el cabello largo y lleva una chaqueta negra. El hombre con el suéter azul lleva una camisa de manga larga y una chaqueta negra. La mujer con la chaqueta negra está de pie junto al hombre con el suéter azul.

# POR QUÉ ES UN PROBLEMA COMPLEJO?



Kahoot!



<https://bit.ly/3QXkbPk>

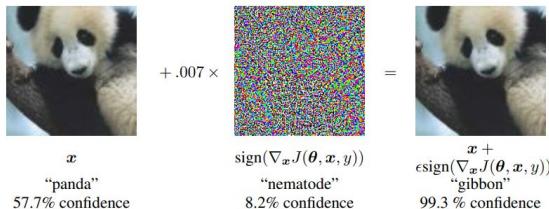
# COMPLEJIDAD DE LA PERCEPCIÓN VISUAL

Published as a conference paper at ICLR 2015

## EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES

Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy  
 Google Inc., Mountain View, CA  
 {goodfellow, shlens, szegedy}@google.com

Published as a conference paper at ICLR 2015



## One Pixel Attack for Fooling Deep Neural Networks

Jiawei Su\*, Danilo Vasconcelos Vargas\* and Kouichi Sakurai

: 2019

*Abstract*—Recent research has revealed that the output of Deep Neural Networks (DNN) can be easily altered by adding relatively small perturbations to the input vector. In this paper, we analyze an attack in an extremely limited scenario where only one pixel can be modified. For that we propose a one-pixel attack for generating one-pixel adversarial examples based on differential evolution (DE). It requires less adversarial information (a black-box attack) and can fool more types of networks due to the



*Abstraer y comprender imágenes de forma robusta es un desafío formidable para las computadoras, lo que resalta la complejidad de replicar la sofisticada capacidad de percepción visual humana mediante algoritmos.*

# PRINCIPALES TAREAS

Logic could be learned from images

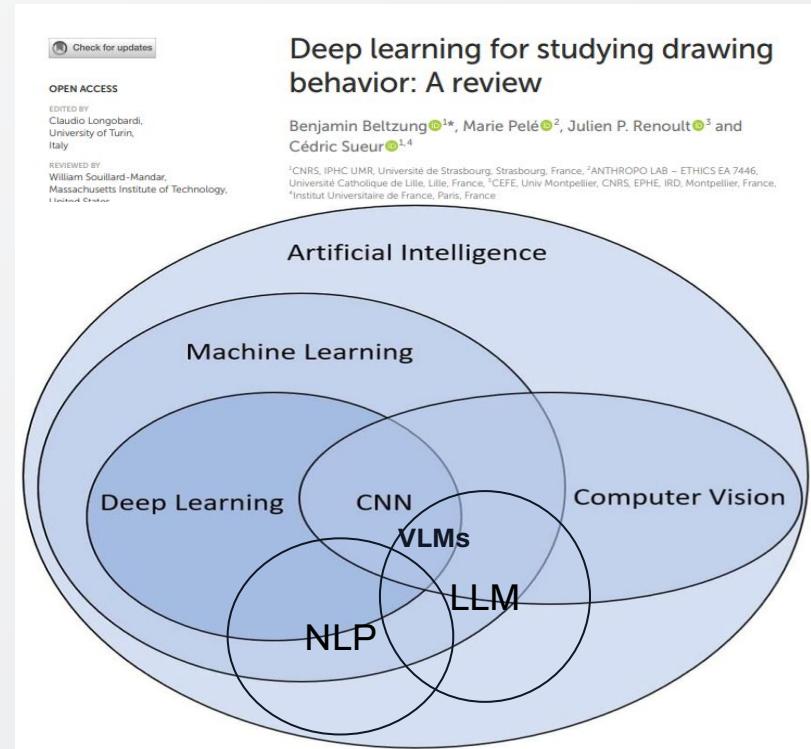
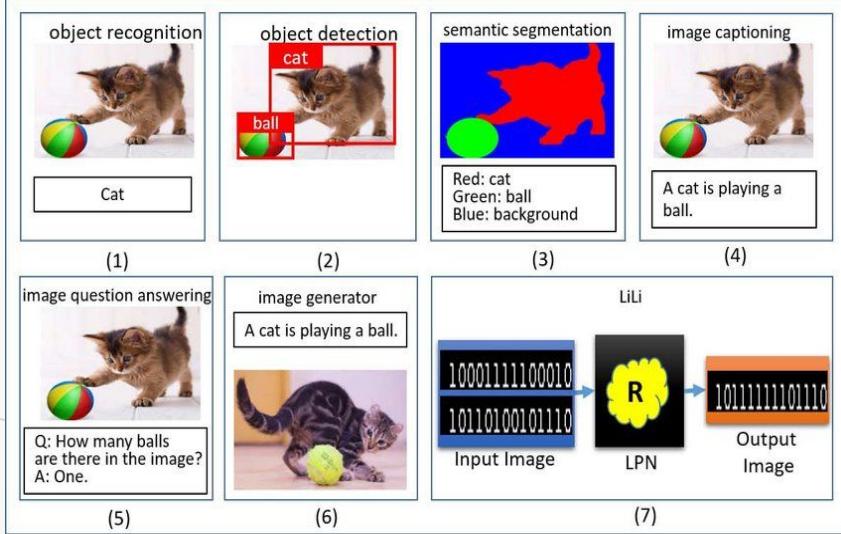
Qian Guo<sup>a,c</sup>, Yuhua Qian <sup>\*a,b,c</sup>, Xinyan Liang<sup>a,c</sup>, Yanhong She<sup>d</sup>, Deyu Li<sup>b,e</sup>, Jiye Liang<sup>b,c</sup>

<sup>a</sup>Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, Shanxi, China

<sup>b</sup>Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, Shanxi, China

<sup>c</sup>School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China

<sup>d</sup>College of Science, Xi'an Shiyou University, Xi'an 710065, Shaan'xi, China



## Deep learning for studying drawing behavior: A review

Benjamin Beltzung<sup>①\*</sup>, Marie Pelé<sup>②</sup>, Julien P. Renoult<sup>③</sup> and

Cédric Sueur<sup>④,1,4</sup>

<sup>①</sup>CNRS, IPHC UMR, Université de Strasbourg, Strasbourg, France, <sup>②</sup>ANTHROPO LAB – ETHICS EA 7446, Université Catholique de Lille, Lille, France, <sup>③</sup>CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France,

<sup>④</sup>Institut Universitaire de France, Paris, France

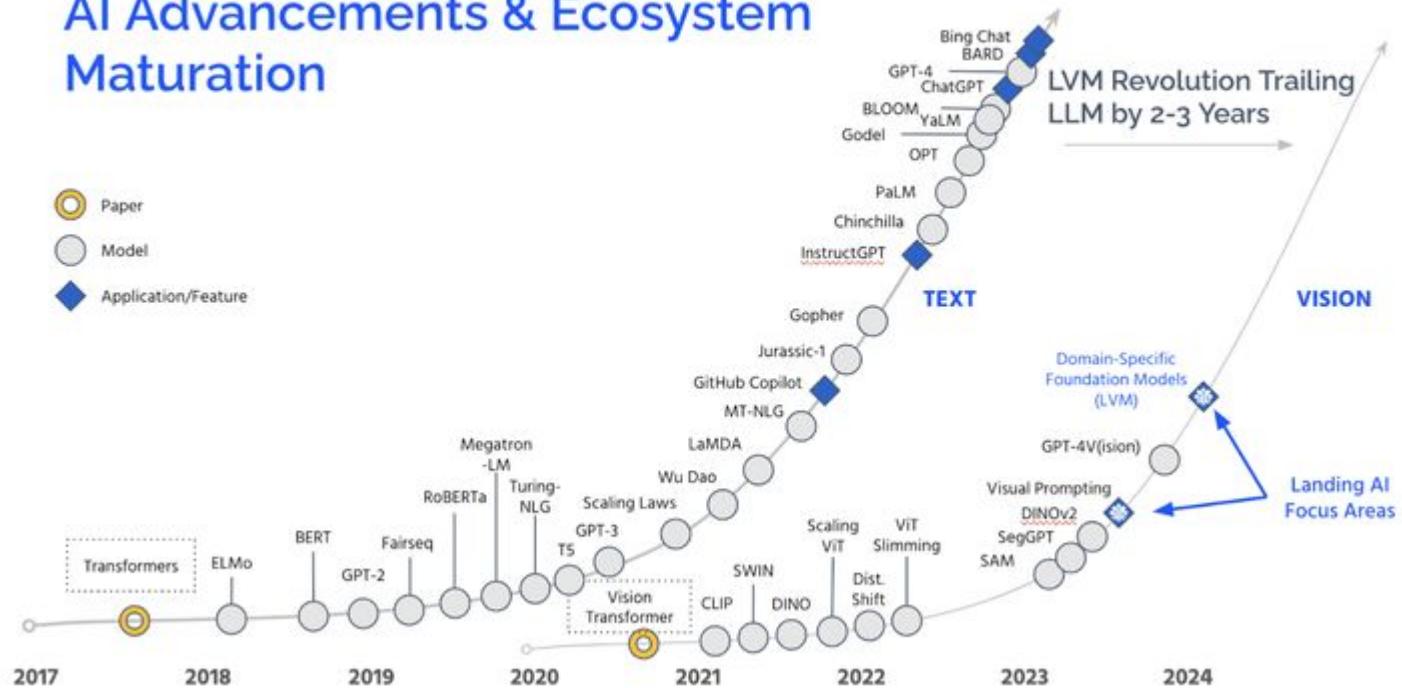
# PRINCIPALES HERRAMIENTAS



# CÓMO HA EVOLUCIONADO?

## AI Advancements & Ecosystem Maturation

- Paper
- Model
- ◆ Application/Feature



02



# Aplicaciones

Cómo y con qué se trabaja  
Computer Vision

# Extracción de información



**Deep learning for OCR**



## Named Entity Recognition (NER)



```
>>> import layoutparser as lp
>>> model = lp.Detection2LayoutModel('lp://PubLayoutNet/mask_rcnn_R_50_FPN_3x/config')
>>> layout = model.detect(image) # You need to load the image
>>> image = layout.image else, image = cv2.cvtColor(..., cv2.COLOR_BGR2GRAY)
>>> lp.draw_box(image, layout, ...) # With extra configurations
```



Figure 1: Annotation Examples in BIDUCNN. (a) and (b) show examples for the labeling of text regions. The boxes are colored differently to reflect the layout element categories. Illustration 1 (a), the sets in which each page view are categorized as title blocks. Illustration 1 (b) shows the same.

Figure 1: Annotation Examples in BIDUCNN. (a) and (b) show examples for the labeling of text regions. The boxes are colored differently to reflect the layout element categories. Illustration 1 (a), the sets in which each page view are categorized as title blocks. Illustration 1 (b) shows the same.

Figure 1: Annotation Examples in BIDUCNN. (a) and (b) show examples for the labeling of text regions. The boxes are colored differently to reflect the layout element categories. Illustration 1 (a), the sets in which each page view are categorized as title blocks. Illustration 1 (b) shows the same.

Figure 1: Annotation Examples in BIDUCNN. (a) and (b) show examples for the labeling of text regions. The boxes are colored differently to reflect the layout element categories. Illustration 1 (a), the sets in which each page view are categorized as title blocks. Illustration 1 (b) shows the same.

**Paper with Complex Layouts**

# LayoutParser

## LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis

Zejiang Shen<sup>1</sup> (✉), Ruochen Zhang<sup>2</sup>, Melissa Dell<sup>3</sup>, Benjamin Charles Germain Lee<sup>4</sup>, Jacob Carlson<sup>3</sup>, and Weining Li<sup>5</sup>

<sup>1</sup> Allen Institute for AI

shannons@allenai.org

<sup>2</sup> Brown University

ruochen\_zhang@brown.edu

<sup>3</sup> Harvard University

{melissadell,jacob.carlson}@fas.harvard.edu

<sup>4</sup> University of Washington

bcl@cs.washington.edu

<sup>5</sup> University of Waterloo

w422li@uwaterloo.ca

**Abstract.** Recent advances in document image analysis (DIA) have been primarily driven by the application of neural networks. Ideally, research outcomes could be easily deployed in production and extended for further investigation. However, various factors like loosely organized codebases and sophisticated model configurations complicate the easy reuse of important innovations by a wide audience. Though there have been on-going efforts to improve reusability and simplify deep learning (DL) model development in disciplines like natural language processing and computer vision, none of them are optimized for challenges in the domain of DIA. This represents a major gap in the existing toolkit, as DIA is central to academic research across a wide range of disciplines in the social sciences and humanities. This paper introduces LayoutParser, an open-source library for streamlining the usage of DL in DIA research and applications. The core LayoutParser library comes with a set of simple and intuitive interfaces for applying and customizing DL models for layout detection, character recognition, and many other document processing tasks. To promote extensibility, LayoutParser also incorporates a community platform for sharing both pre-trained models and full document digitization pipelines. We demonstrate that LayoutParser is helpful for both lightweight and large-scale digitization pipelines in real-word use cases. The library is publicly available at <https://layout-parser.github.io>.

**Keywords:** Document Image Analysis · Deep Learning · Layout Analysis · Character Recognition · Open Source library · Toolkit.

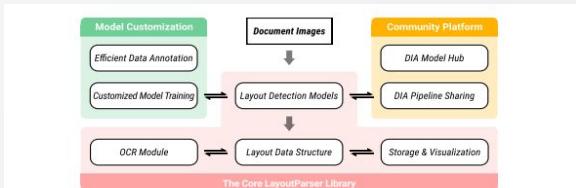


Fig. 1: The overall architecture of LayoutParser. For an input document image, the core LayoutParser library provides a set of off-the-shelf tools for layout detection, OCR, visualization, and storage, backed by a carefully designed layout data structure. LayoutParser also supports high level customization via efficient layout annotation and model training functions. These improve model accuracy on the target samples. The community platform enables the easy sharing of DIA models and whole digitization pipelines to promote reusability and reproducibility. A collection of detailed documentation, tutorials and exemplar projects make LayoutParser easy to learn and use.

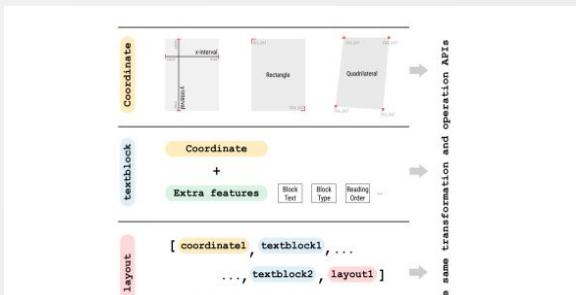
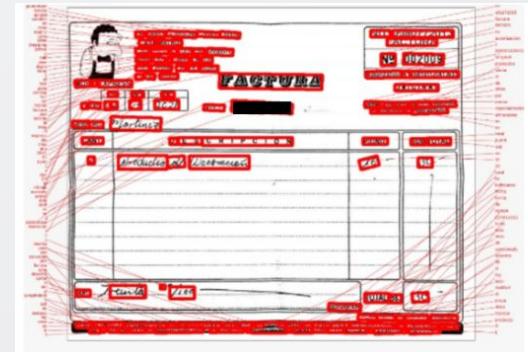
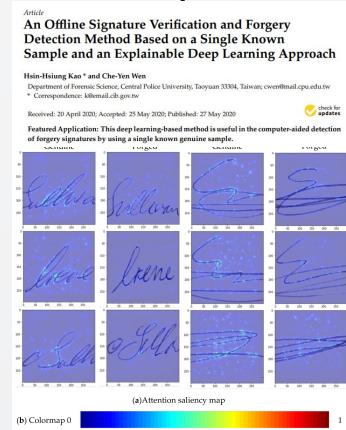


Fig. 2: The relationship between the three types of layout data structures. Coordinate supports three kinds of variation; TextBlock consists of the coordinate information and extra features like block text, types, and reading orders; a Layout object is a list of all possible layout elements, including other Layout objects. They all support the same set of transformation and operation APIs for maximum flexibility.

# Validación de documentos



# Reniec BioFacial

Face landmark detection

 gob.pe | Plataforma digital única del Estado Peruano

[Inicio](#) > [El Estado](#) > [RENIEC](#) > [Noticias](#) > RENIEC lanza aplicativo DNI BioFacial para que ciudadanos tramiten su documento desde celulares iOS

**Registro Nacional de Identificación y Estado Civil**

**RENIEC lanza aplicativo DNI BioFacial para que ciudadanos tramiten su documento desde celulares iOS**

Nota de prensa  
5 de noviembre de 2021 - 6:27 a. m.

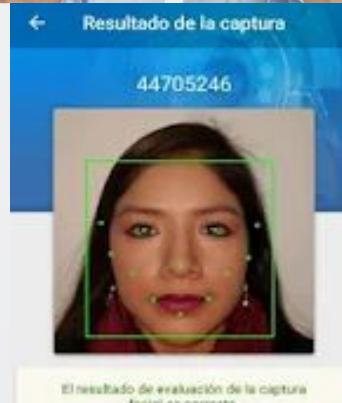
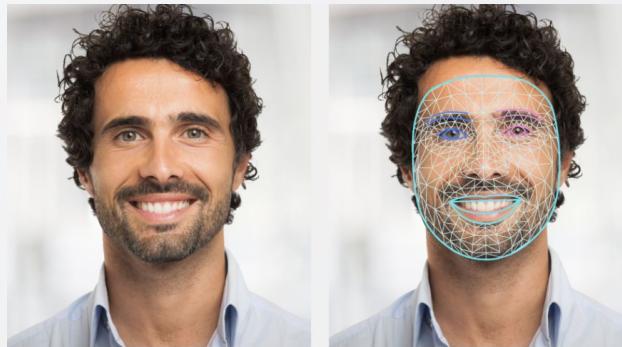
Con el fin de que la población realice sus trámites sin la necesidad de salir de casa, el Registro Nacional de Identificación y Estado Civil - RENIEC puso a disposición de la ciudadanía, la versión del aplicativo DNI BioFacial para celulares con sistema operativo iOS.

El DNI BioFacial puede descargarse del Play Store de Android y solamente requiere como requisito que el celular tenga una memoria RAM superior a 3 GB, en el caso de iOS, puede descargarse del APP STORE de APPLE.

El aplicativo, ya se puede descargar sin costo alguno, desde el App Store de APPLE y en la misma línea de la versión para celulares con el sistema operativo Android, permite realizar trámites de duplicado/rectificación de domicilio, renovación por caducidad del DNI y Registro Digital de Nacimiento.

El aplicativo permite la toma de fotografía para realizar la verificación biométrica facial de la identidad de la persona que realiza alguno de los trámites antes mencionados, para luego completar el proceso a través de la página web del RENIEC.

Para realizar la toma de la fotografía con el aplicativo DNI BioFacial, se debe contar con el apoyo de otra persona, quien debe colocar el celular aproximadamente a unos 50 centímetros del rostro, de manera tal que ambos queden en posición paralela entre sí y perpendicular con respecto al suelo.



# Seguridad y salud

Research Article  
**Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks**

Yange Li,<sup>1</sup> Han Wei,<sup>1</sup> Zheng Han<sup>2</sup>,<sup>✉</sup> Jianling Huang,<sup>1</sup> and Weidong Wang<sup>1,2</sup>

<sup>1</sup>School of Civil Engineering, Central South University, Changsha 410075, China

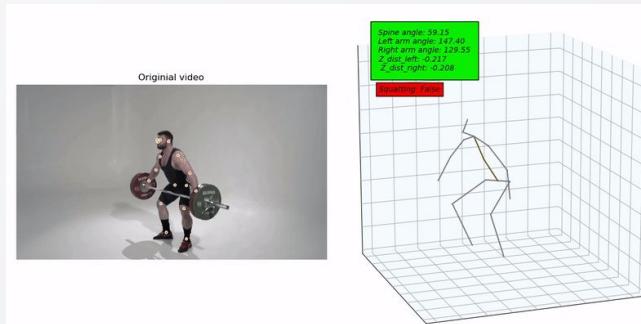
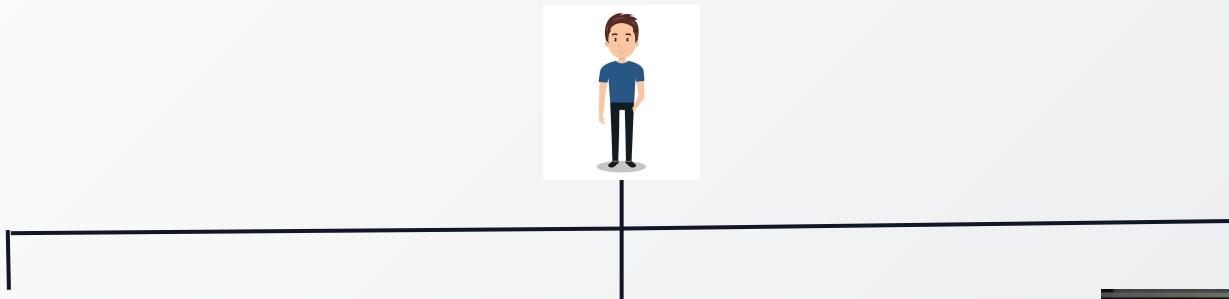
<sup>2</sup>The Key Laboratory of Engineering Structures of Heavy Haul Railway, Ministry of Education, Changsha 410075, China

Correspondence should be addressed to Zheng Han, zheng\_han@csu.edu.cn

Received 8 August 2019; Revised 14 May 2020; Accepted 10 September 2020; Published 19 September 2020



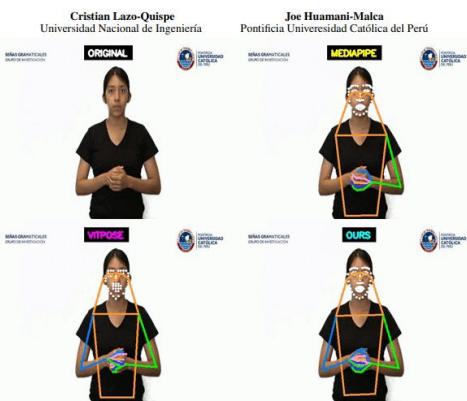
FIGURE 9. Validation results.



# Impacto social



## Impact of Pose Estimation Models for Landmark-based Sign Language Recognition



## Article Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks

Hasan Alkahtani<sup>1,2</sup>, Theyazn H. H. Aldhyani<sup>1,3,\*</sup> and Mohammed Y. Alzahrani<sup>1,4</sup>

- <sup>1</sup> King Salman Center for Disability Research, Riyadh 11614, Saudi Arabia; hasalkahhtani@kfu.edu.sa (H.A.)
  - <sup>2</sup> Applied Computer Sciences and Information Technology, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia
  - <sup>3</sup> Applied College in Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia
  - <sup>4</sup> Department of Computer Sciences and Information Technology, Al-Baha University, Al-Baha 65799, Saudi Arabia
- \*Correspondence: talidhyani@kfu.edu.sa

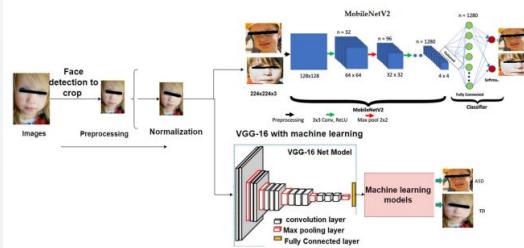
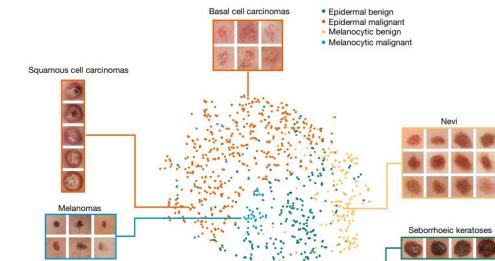


Figure 3. Framework of the proposed recognition system.

## Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva<sup>1\*</sup>, Brett Kuprel<sup>1\*</sup>, Roberto A. Novoa<sup>2,3</sup>, Justin Ko<sup>3</sup>, Susan M. Swetter<sup>3,4</sup>, Helen M. Blau<sup>3</sup> & Sebastian Thrun<sup>5</sup>



# Sign Language Recognition

## Impact of Pose Estimation Models for Landmark-based Sign Language Recognition

**Cristian Lazo-Quispe**  
 Universidad Nacional de Ingeniería  
 Lima, Perú  
 clazoq@uni.pe

**Joe Huamani-Malca**  
 Pontificia Universidad Católica del Perú  
 Lima, Perú  
 huamani.jn@pucp.edu.pe

**Manuel Stev Harold Huamán-Ramos**  
 Pontificia Universidad Católica del Perú  
 Lima, Perú  
 mhuamanr@pucp.edu.pe

**Gissella Bejarano**  
 Baylor University  
 Texas, USA  
 gissella\_bejaranonc@baylor.edu

**Pablo Rivas**  
 Baylor University  
 Texas, USA  
 pablo\_rivas@baylor.edu

**Tomas Cerny**  
 Baylor University  
 Texas, USA  
 tomas\_cerny@baylor.edu

### Abstract

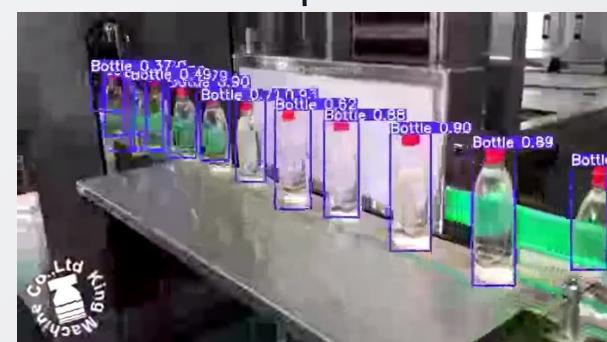
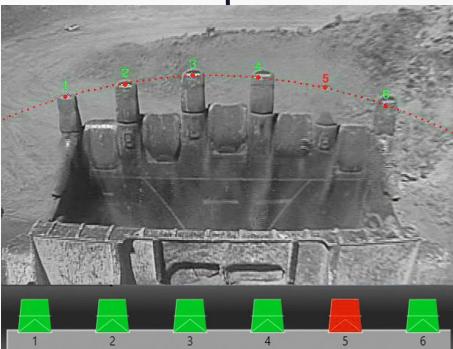
Sign Language Recognition (SLR) models rely heavily on advances reached by Human Action Recognition (HAR). One of the simplest and most dimensional-reduced modalities is the skeleton joints and limbs represented with key-point landmarks and edges connecting these landmarks. These skeletons can be obtained by pose estimation, depth maps, or motion capture. For HAR, models are usually interested in less granularity of pose estimation compared to SLR, where the landmark estimation of not only the pose and body but the facial gestures, hands, and fingers is crucial. In this work, we compare three whole-body estimation libraries/models that are gaining attraction in the SLR task. We first find their relation by identifying common keypoints in their landmark structure and analyzing their quality. Then, we complement this analysis by comparing their annotations in three sign language datasets with videos of different quality, backgrounds, and regions (Peru and USA). Finally, we test a sign language recognition model to compare the quality of the annotations provided by these libraries/models.

SLR Model	Library	Top-5					
		AEC		PUCP		WLAL	
		29	71	29	71	29	71
Spoter	MediaPipe	0.869 ± 0.0125	0.882 ± 0.0134	0.751 ± 0.032	0.735 ± 0.015	<b>0.893</b> ± 0.010	<b>0.919</b> ± 0.008
	OpenPose	0.879 ± 0.010	0.869 ± 0.024	0.784 ± 0.024	0.777 ± 0.013	0.797 ± 0.007	0.838 ± 0.012
	WholePose	<b>0.8854</b> ± 0.014	<b>0.905</b> ± 0.015	0.781 ± 0.016	0.758 ± 0.016	0.764 ± 0.030	0.801 ± 0.015
	SmileLab	0.813 ± 0.007	0.803 ± 0.021	0.608 ± 0.040	0.613 ± 0.028	<b>0.876</b> ± 0.012	<b>0.796</b> ± 0.022
SmileLab	OpenPose	0.842 ± 0.011	0.853 ± 0.015	0.700 ± 0.027	<b>0.697</b> ± 0.030	0.798 ± 0.019	0.734 ± 0.018
	WholePose	<b>0.866</b> ± 0.007	<b>0.854</b> ± 0.015	0.690 ± 0.044	0.689 ± 0.019	0.812 ± 0.024026	0.767 ± 0.015

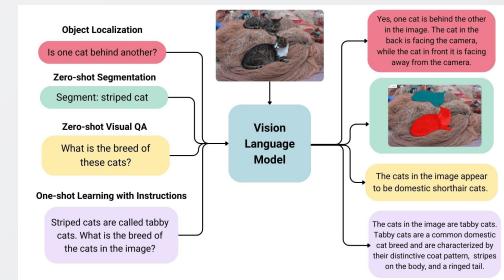
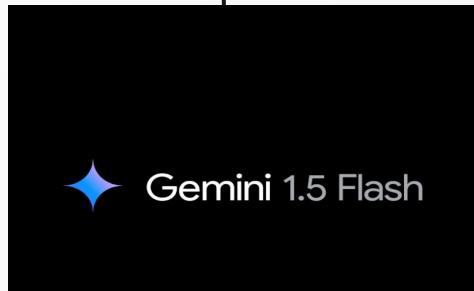
Table 2: Spoter [5] and SmileLab [14] Top-5 results for groups of keypoints: 29 and 71



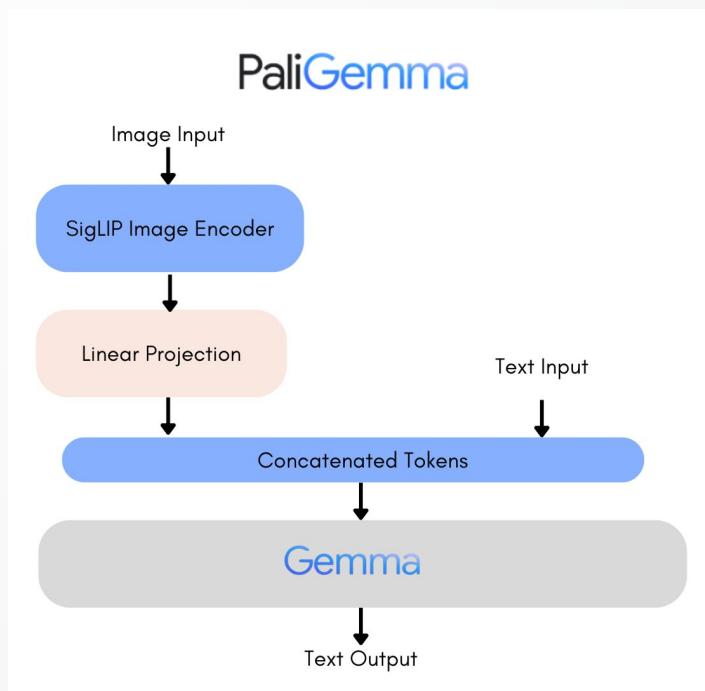
# Otros sectores



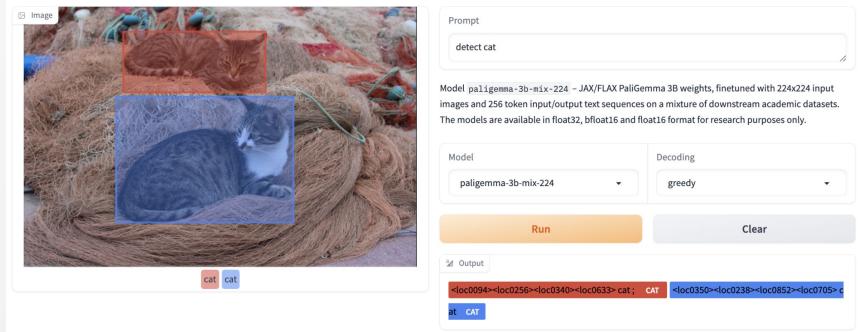
# Vision Language Models



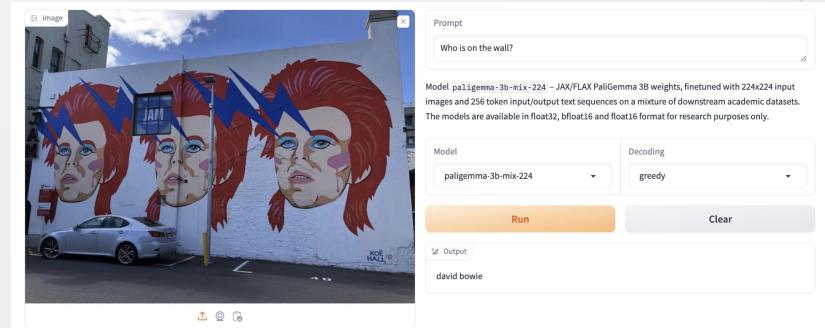
# PaliGemma



## Detection



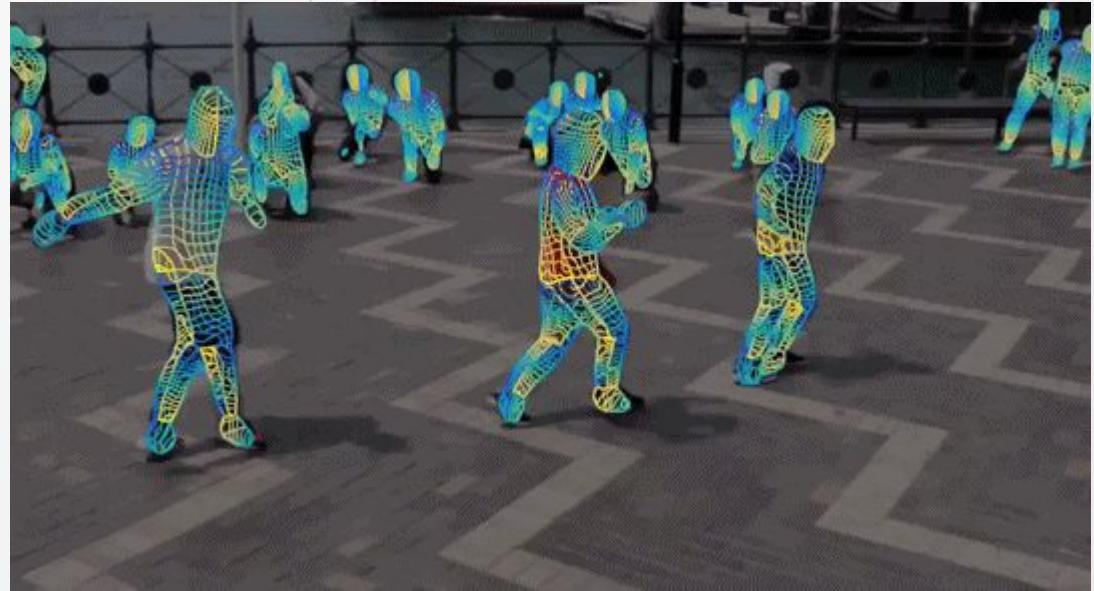
## Visual Question Answering



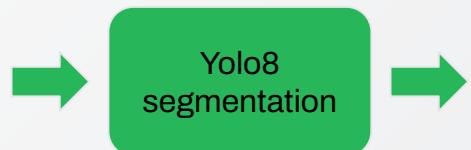
03

# Demo

¡Código en vivo!

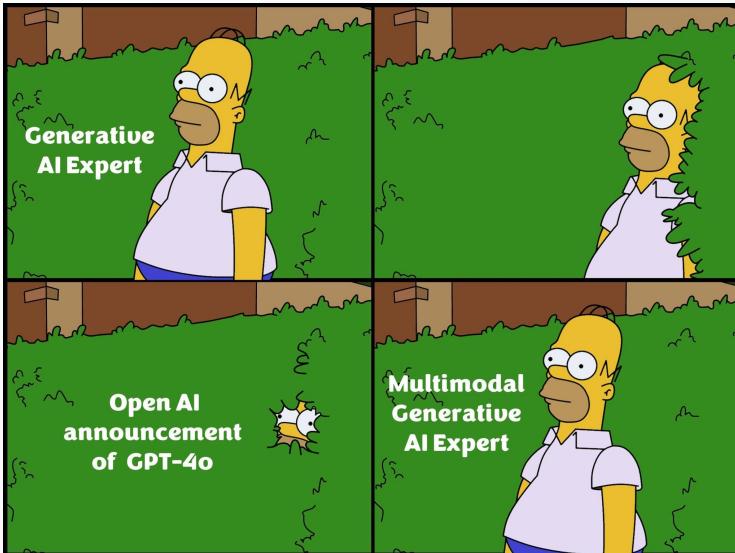


# Demo 1: Object segmentation



<https://bit.ly/3KfvU88>

# Demo 2: Visual Question Answering



are they in an  
office?



ViLT:  
Vision-and-Language  
Transformer

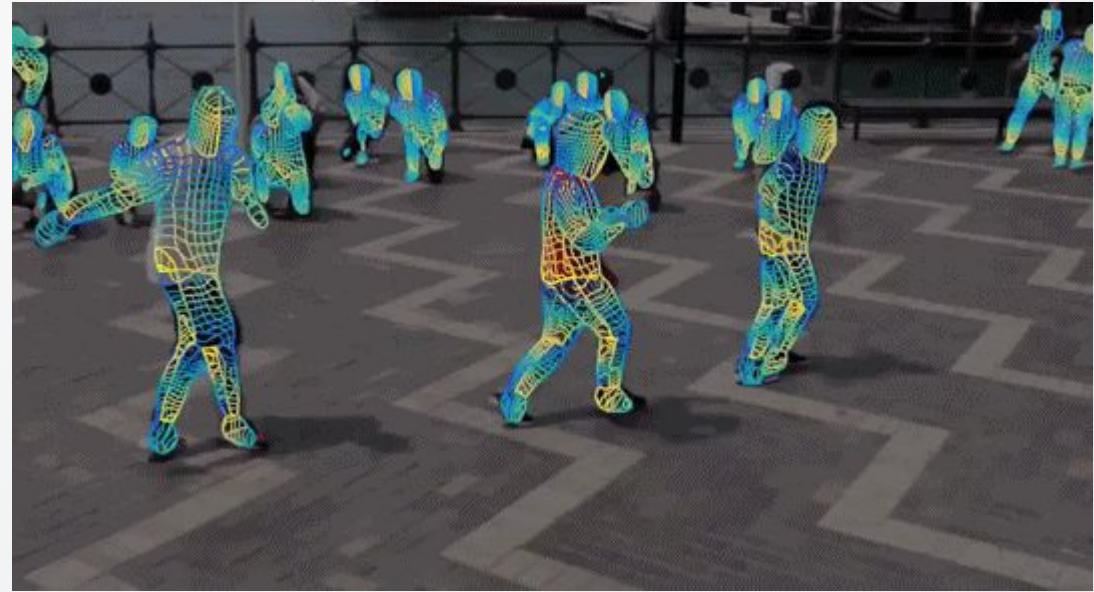
→ Respuesta

<https://bit.ly/3RkfoYL>

04

# Ética

Discusió sobre el futur



# Sesgos en modelos

2015

**Forbes**

## Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software

Maggie Zhang Forbes Staff  
I write about technology, innovation, and startups.

Follow

JUL 1, 2015, 01:42pm EDT

This article is more than 8 years old.



The article discusses a controversial incident where Google Photos mislabeled two African-American men as gorillas due to a facial recognition error. The image shows a man speaking on stage at a Google Photos event.

<https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software>

2021

**The New York Times**

Artificial Intelligence > | Harvesting Data for A.I. | A.I. Data Race, Explained | What Is 'Synthetic Data'? | Key Figures in the Field | A.I. Faces Quiz

## Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men

Facebook called it "an unacceptable error." The company has struggled with other issues related to race.

Share this article



A video thumbnail from Facebook showing a screen with a thumbs-up icon and the text "Facebook Way". The video is captioned "Facebook apologized on Friday for mislabeling and said it was looking into its recommendation feature to 'prevent this from happening again.' Jon Wilson/The New York Times

By Ryan Mcc  
Published Sept. 3, 2021 Updated Oct. 4, 2021

59.9K 11.6K Comments 17.8K Shares

Like Comment Share

Daily Mail • Follow Jun 27, 2020 White man calls cops on black men at marina



A video thumbnail from Daily Mail showing a man in a green shirt talking to another man. The caption reads "Where is the white shirt?"

Keep seeing videos about Primates? Yes Dismiss

<https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html>

# ¿Estás a favor o en contra del uso de IA para monitorear a los niños en el colegio?



WSJ

# ¿Estás a favor o en contra del uso de IA para monitorear a los niños en el colegio?

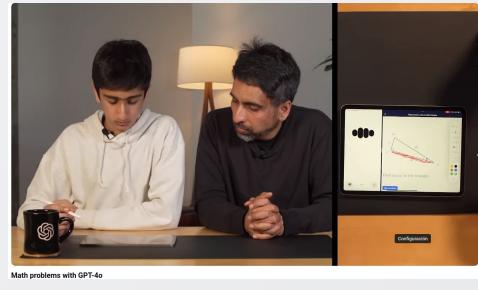
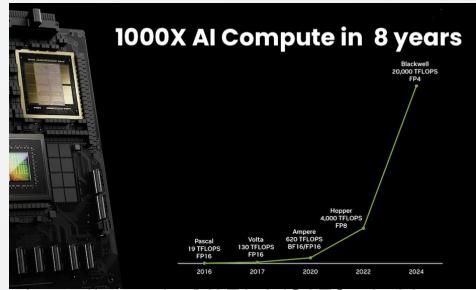
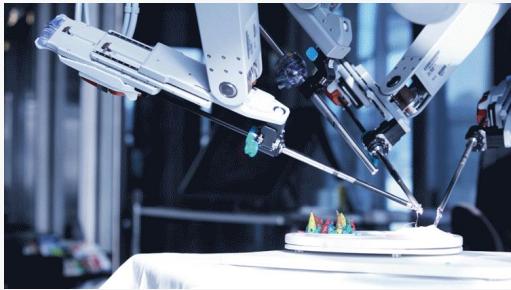
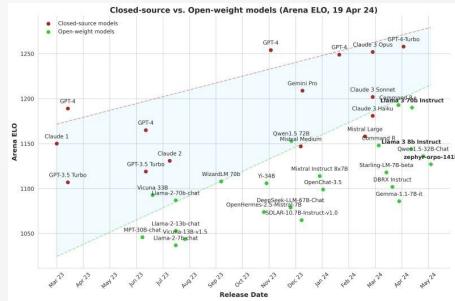
Ficción



Realidad



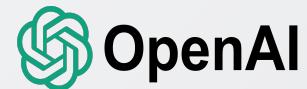
# Avances y tendencias actuales



# ¿Quién mandará en la inteligencia artificial?



DeepMind



OpenAI



cohere



Hugging Face



NVIDIA®

**Muchas gracias!**