

Data Mining en Economía y Finanzas

Universidad de Buenos Aires
Maestría en Data Mining
2021



Historia reciente de la Ciencia de Datos

The Breakthrough

2012 Large Scale Visual Recognition Challenge

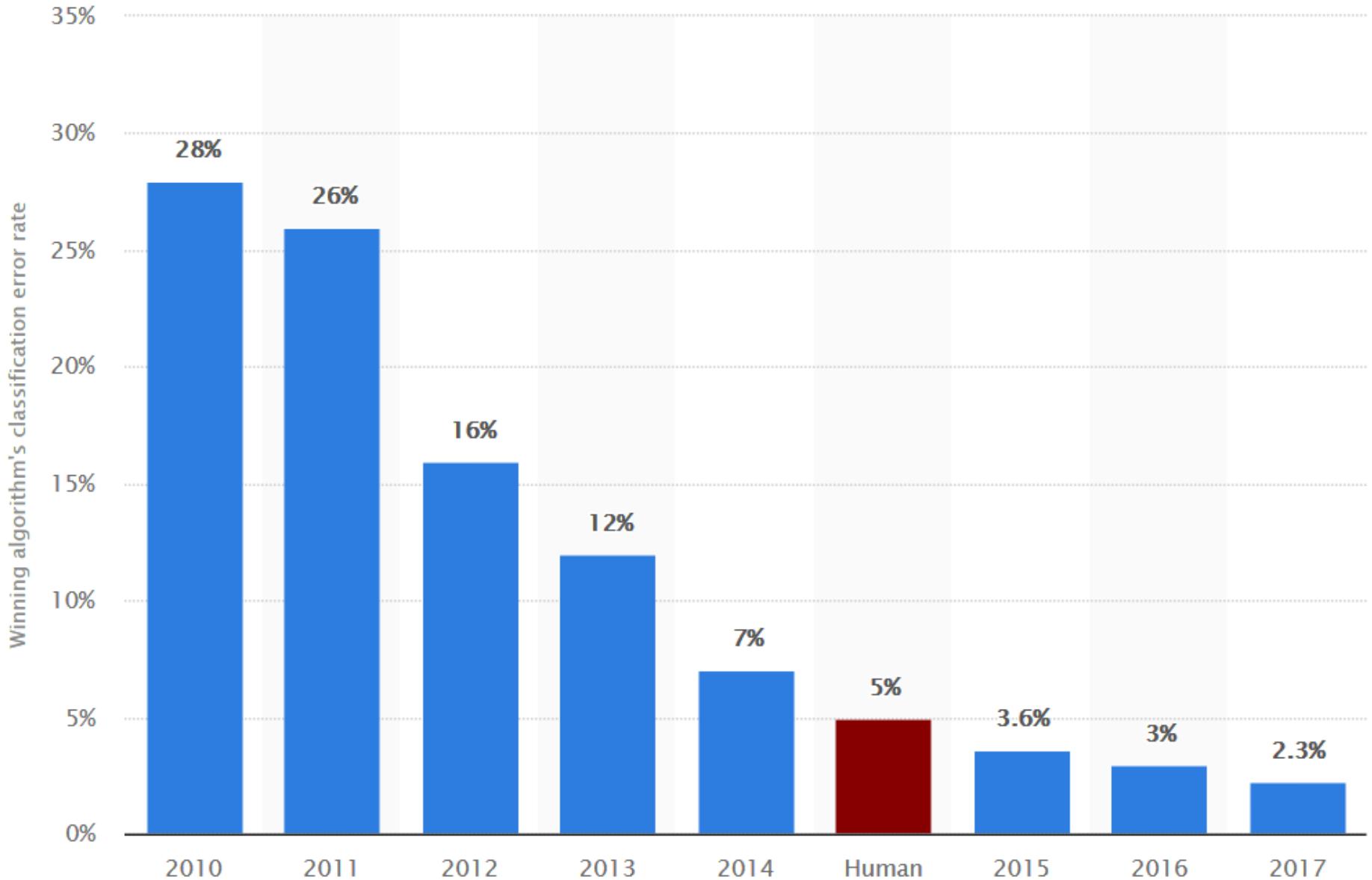
based on Imagenet repository containing millions of
labeled images built in 2010

Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton
presented a solution based on
GPUs x50 speedup

dropped the error rate from 26% to 16%

Supervised Learning

ILSVRC Error Rate Evolution



ImageNet 14M labeled images

Classification



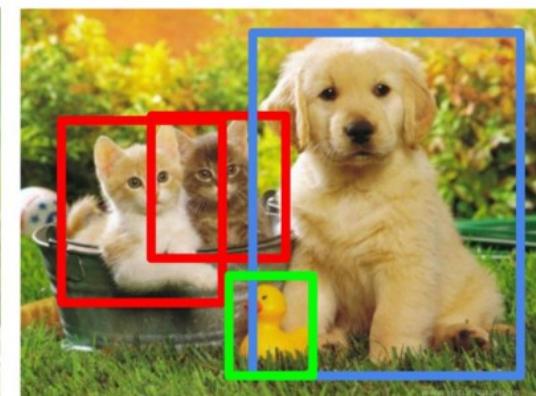
CAT

Classification + Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation



CAT, DOG, DUCK

Single object

Multiple objects

BIG DATA



Pedestrian Collision Warning
at daytime and night-time



Recognition

Lane Departure

year 2013, 1 week to train a DNN

GOOGLE DATACENTER



1,000 CPU Servers
2,000 CPUs • 16,000 cores

600 kWatts
\$5,000,000

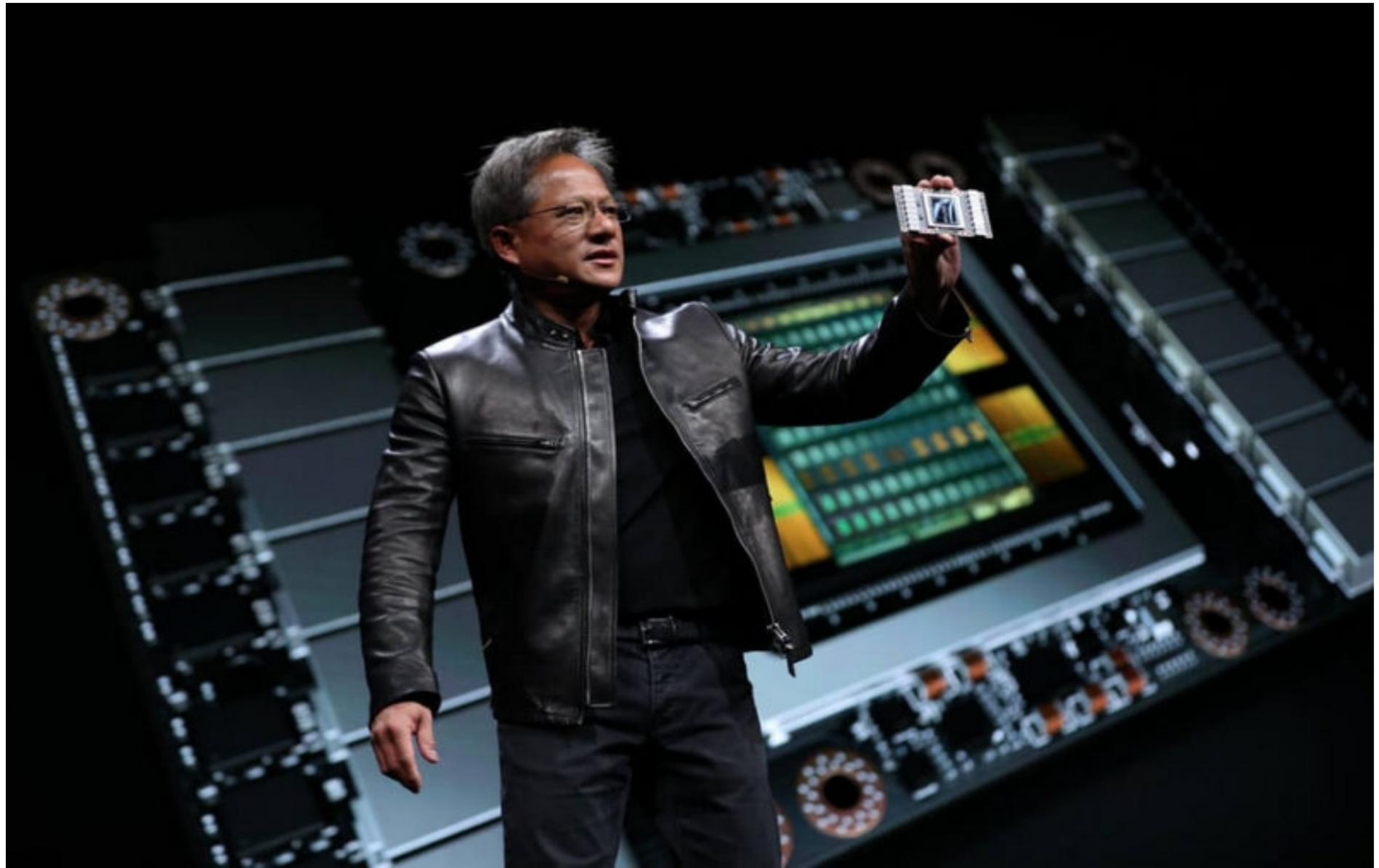
STANFORD AI LAB



3 GPU-Accelerated Servers
12 GPUs • 18,432 cores

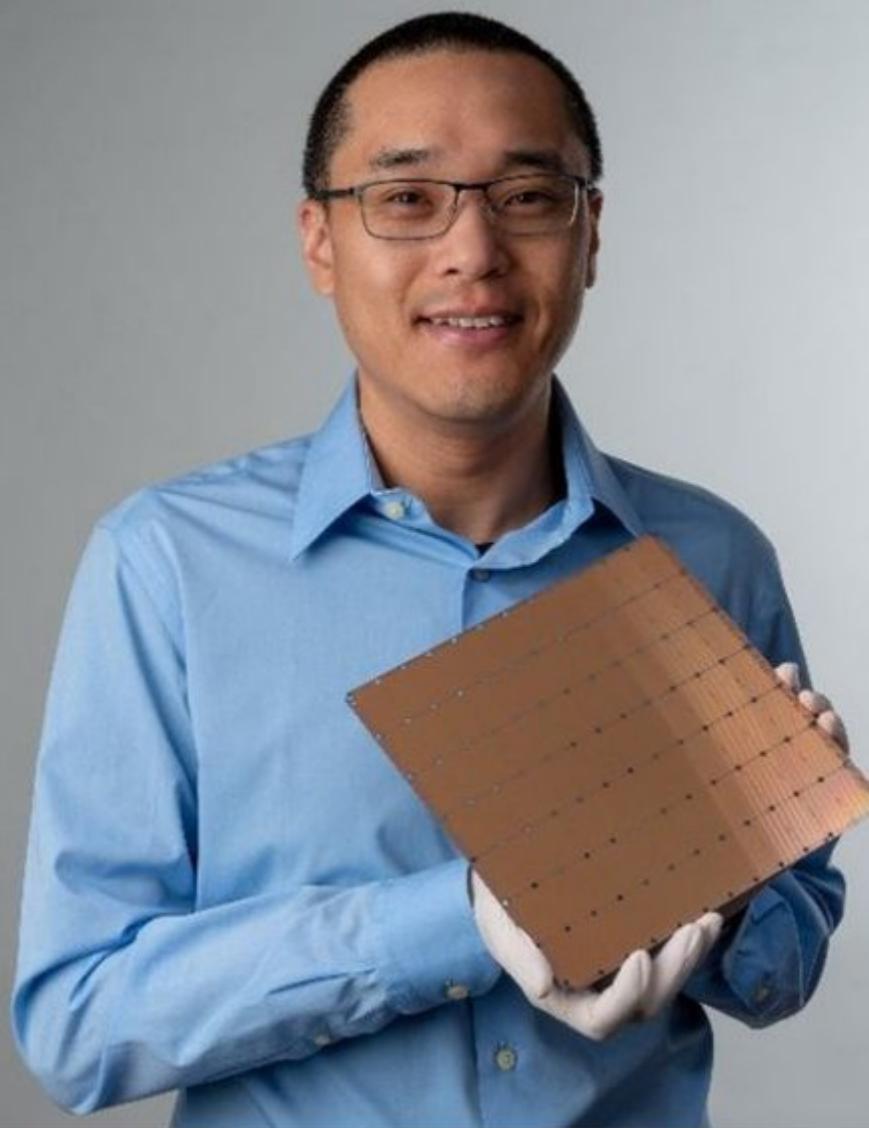
4 kWatts
\$33,000

NVIDIA Tesla V100 USD 8,800



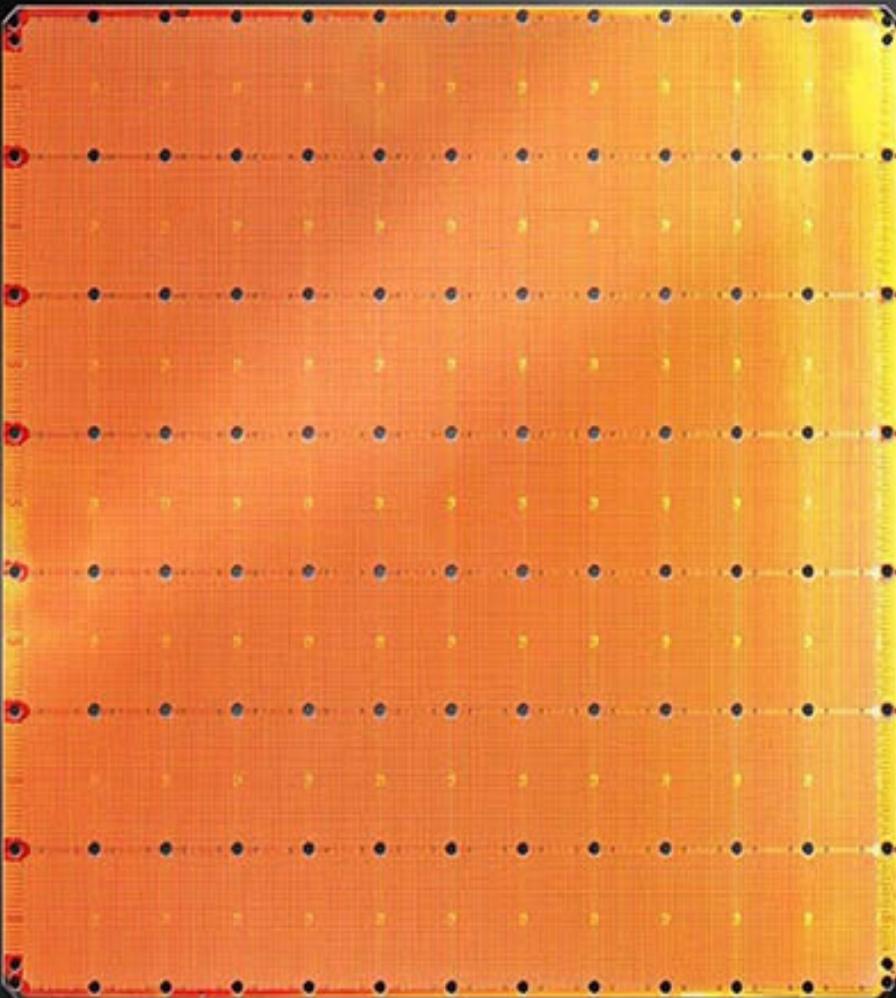
Jensen Huang

accelerating Machine Learning means new hardware. Cerebras



BBC NEWS

Cerebras Wafer Scale Engine

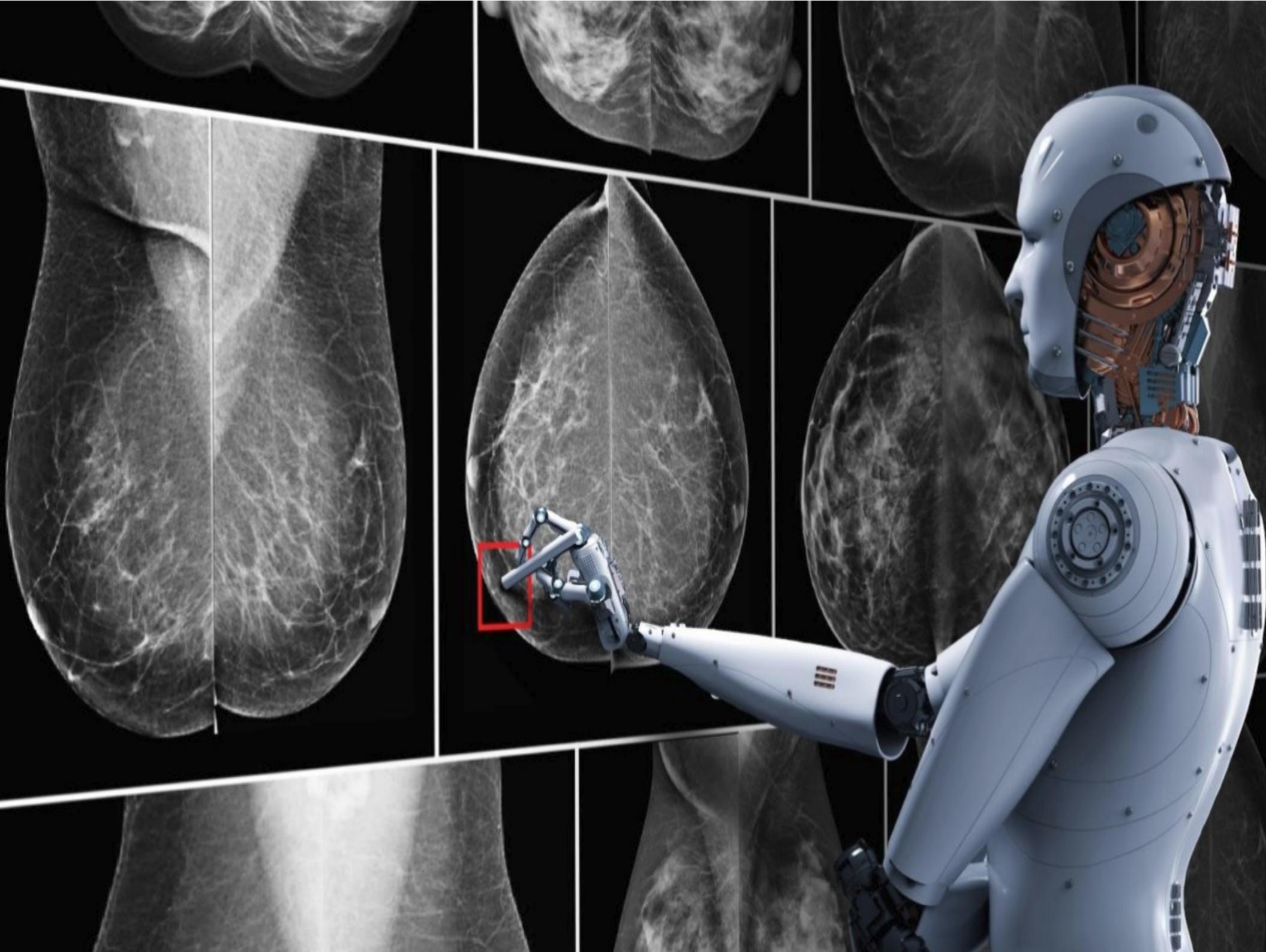


Cerebras WSE
1.2 Trillion Transistors
46,225 mm² Silicon



Largest GPU
21.1 Billion Transistors
815 mm² Silicon

Medicine



The Digital Mammography DREAM Challenge



Out of every 1000 women screened, only 5 will have breast cancer. But 100 will be recalled for further testing.

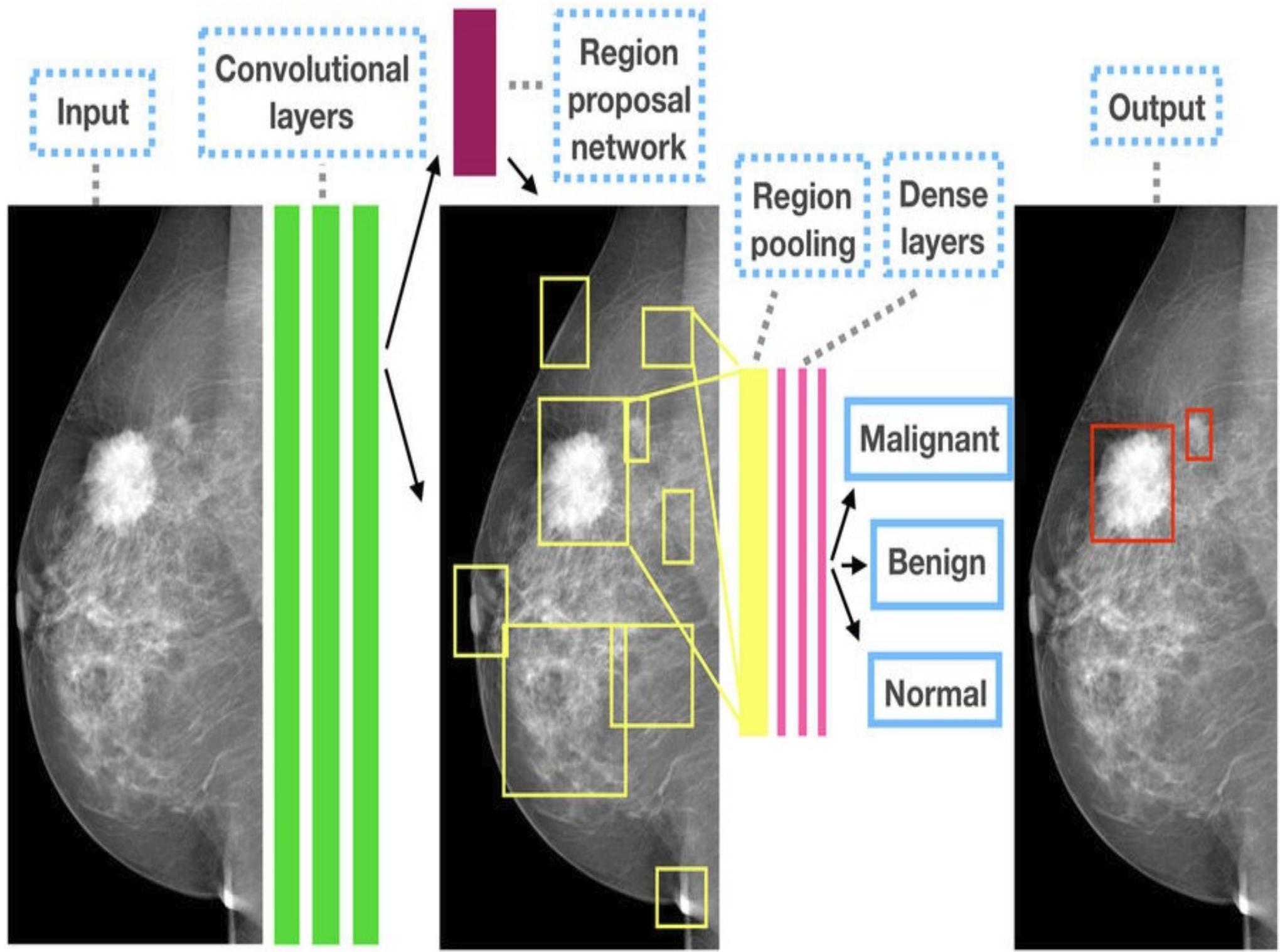
We can do better.

Build a model to help reduce the recall rate for breast cancer screening.

Calling all coders to join the Challenge.

Up to a **\$1,000,000** in cash prizes for winning models.

May the best model win.



Key Points

[JAMA Netw Open](#). 2020 Mar; 3(3): e200265.

Published online 2020 Mar 2. doi: [10.1001/jamanetworkopen.2020.0265](https://doi.org/10.1001/jamanetworkopen.2020.0265)

Question

How do deep learning algorithms perform compared with radiologists in screening mammography interpretation?

Findings

In this diagnostic accuracy study using 144 231 screening mammograms from 85 580 women from the United States and 166 578 screening mammograms from 68 008 women from Sweden, no single artificial intelligence algorithm outperformed US community radiologist benchmarks; including clinical data and prior mammograms did not improve artificial intelligence performance. However, combining best-performing artificial intelligence algorithms with single-radiologist assessment demonstrated increased specificity.

Meaning

Integrating artificial intelligence to mammography interpretation in single-radiologist settings could yield significant performance improvements, with the potential to reduce health care system expenditures and address resource scarcity experienced in population-based screening programs.

Mamografia 3D de 2 a 100 imágenes



Radiologist Salary in the United States

 Search More Jobs

How much does a Radiologist make in the United States? The average Radiologist salary in the United States is: can vary widely depending on many important factors, including [education](#), certifications, additional skills, the number of years of experience, and location. On this website, Salary.com helps you determine your exact pay target.

Recommended Part-time Jobs during Coronavirus Outbreak

[Salary](#) [Salary + Bonus](#) [Benefits](#)

Paid Annually

[View as table](#)

10%

\$311,344

25%

\$360,690

50%(Median)

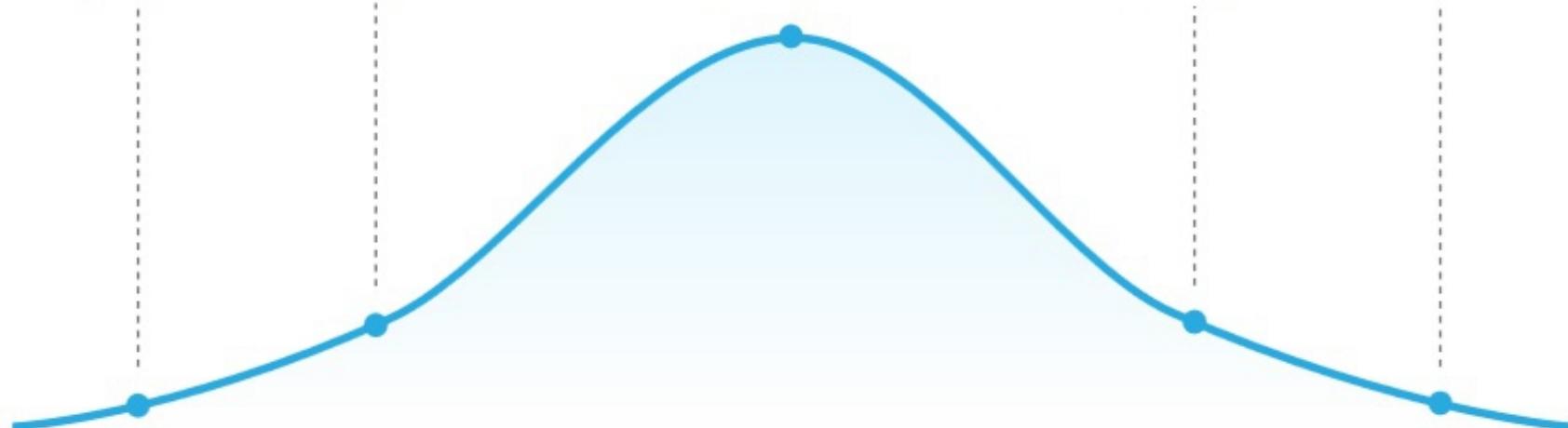
\$414,890

75%

\$479,890

90%

\$539,069





A joint meeting with
 **RCR**
The Royal College of Radiologists

23
MARCH
2018



Bronze sponsors



EverlightRadiology

MEDICA
GROUP

Silver sponsors



ARTIFICIAL INTELLIGENCE IN IMAGING -
THREAT OR OPPORTUNITY?

Venue: etc. venues St Paul's, London

CPD: 6 CREDITS

Researchers worry that AI could be scaring students away from radiology

Marty Stempniak | November 13, 2019 | [Artificial Intelligence](#)



Uncertainty around how AI will reshape the radiology profession may be steering medical students away from it.

That's according to a new survey of healthcare stakeholders, [highlighted](#) in November's *European Journal of Radiology*. Polling med students, radiologists and surgeons about their perceptions of potential threats to the imaging profession, Swiss researchers found that trainees, in particular, expressed worries about the prospect of how AI may reshape the field.

Out of about 55 students polled on these threats, 26% cited artificial intelligence as why they're avoiding imaging as their specialty. Their views stood in contrast to those of radiologists, who see machine learning in a more optimistic light.

"Our survey suggests that medical students might be afraid to specialize in radiology because, among other reasons, they seem to fear the unknown future of AI," Jasper van Hoek, with the University Hospital and University of Bern, Switzerland, and colleagues wrote.

Todos estos
gigantescos avances
han generado
Una

Fiebre por la Ciencia de Datos



Artwork: Tamar Cohen, Andrew J Bu
on a page from a high school yearbo

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2012 Issue

Summary Save Share Comment (18) Print \$8.95 Buy Copies

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected.

WHAT TO READ NEXT



What Data Scientists Re
According to 35 Data Sc

VIEW MORE FROM THE
October 2012 Issue





AMÉRICA TELESHOW TENDENCIAS MIX5411 GRANDES LIBROS
CRIPTO247

TECNO

Científico de datos: por qué es la profesión más sexy del siglo XXI

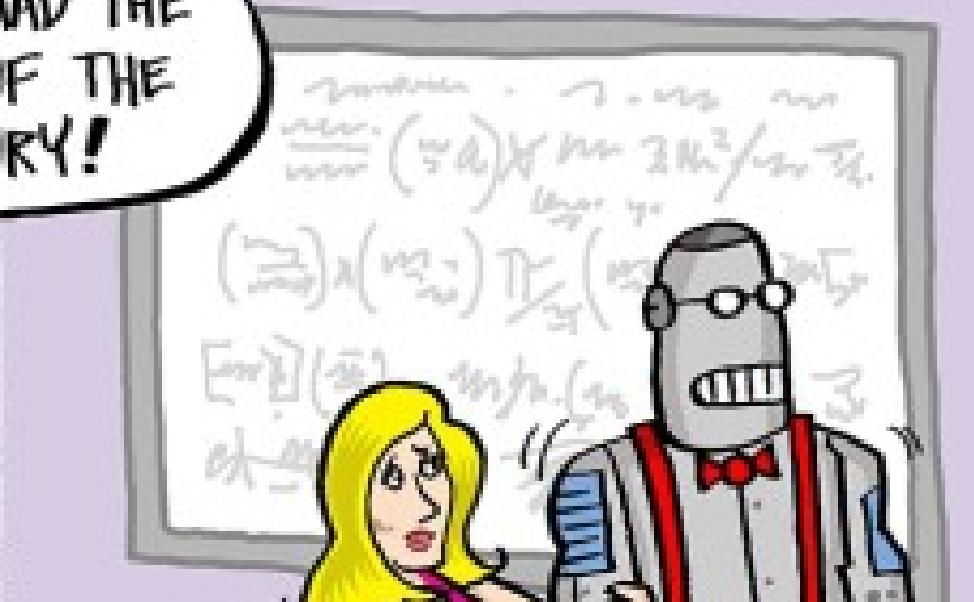
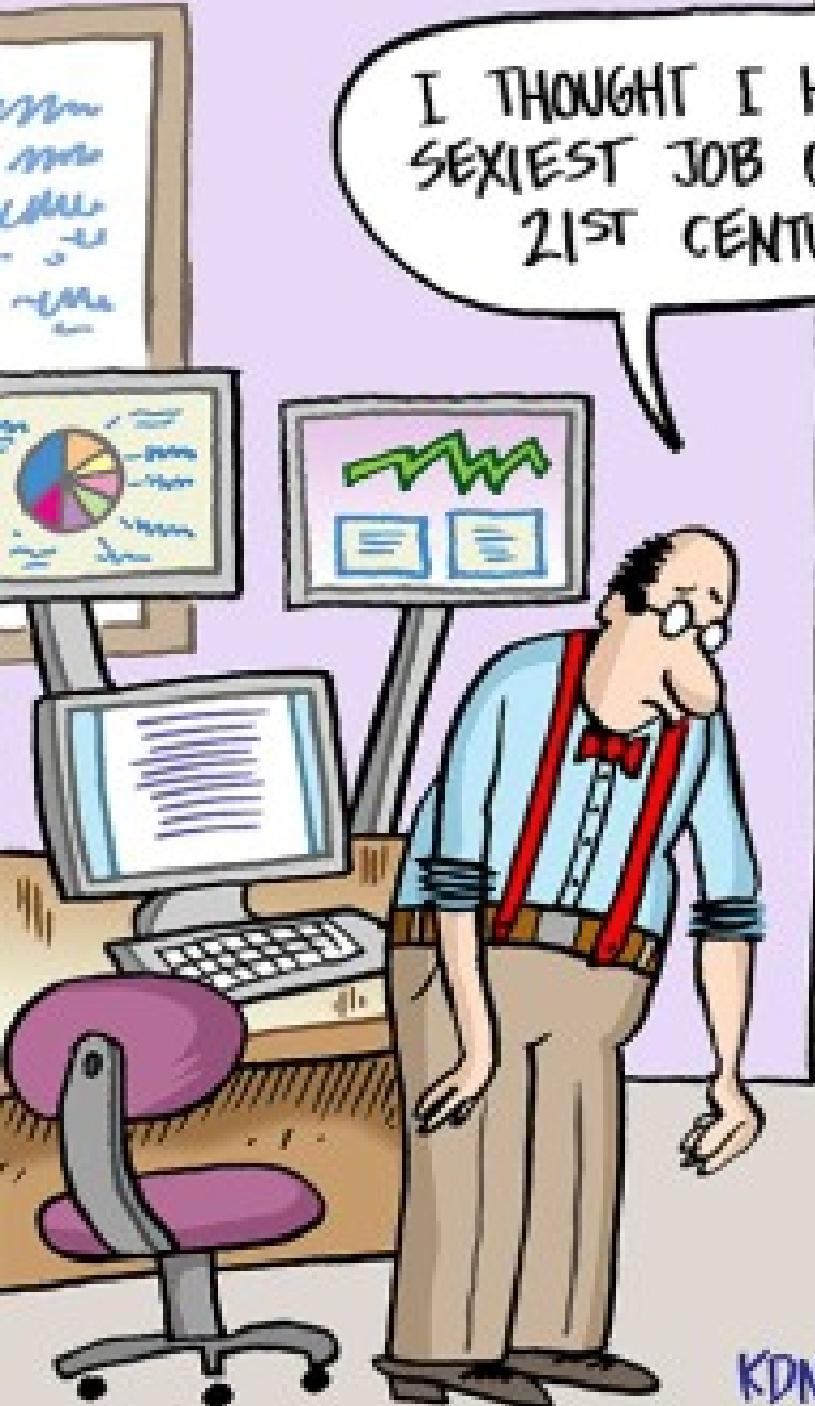
En el mundo de la tecnología es el perfil que más se comenzó a demandar en el último tiempo. Cómo impacta el crecimiento de la automatización en este empleo y en qué tipo de industrias se los necesita



Por Desirée Jaimovich | 9 de mayo de 2017

djaimovich@infobae.com

I THOUGHT I HAD THE
SEXIEST JOB OF THE
21ST CENTURY!



La UBA lanzó una carrera con gran demanda laboral y sueldos de hasta 150 mil pesos

La Facultad de Ciencias Exactas creó la Licenciatura en Ciencias de Datos, un perfil muy requerido por las empresas. De qué se trata la propuesta

16 de Julio de 2020

 Compartir en Facebook

 Compartir en Twitter

MÁS LEÍDAS

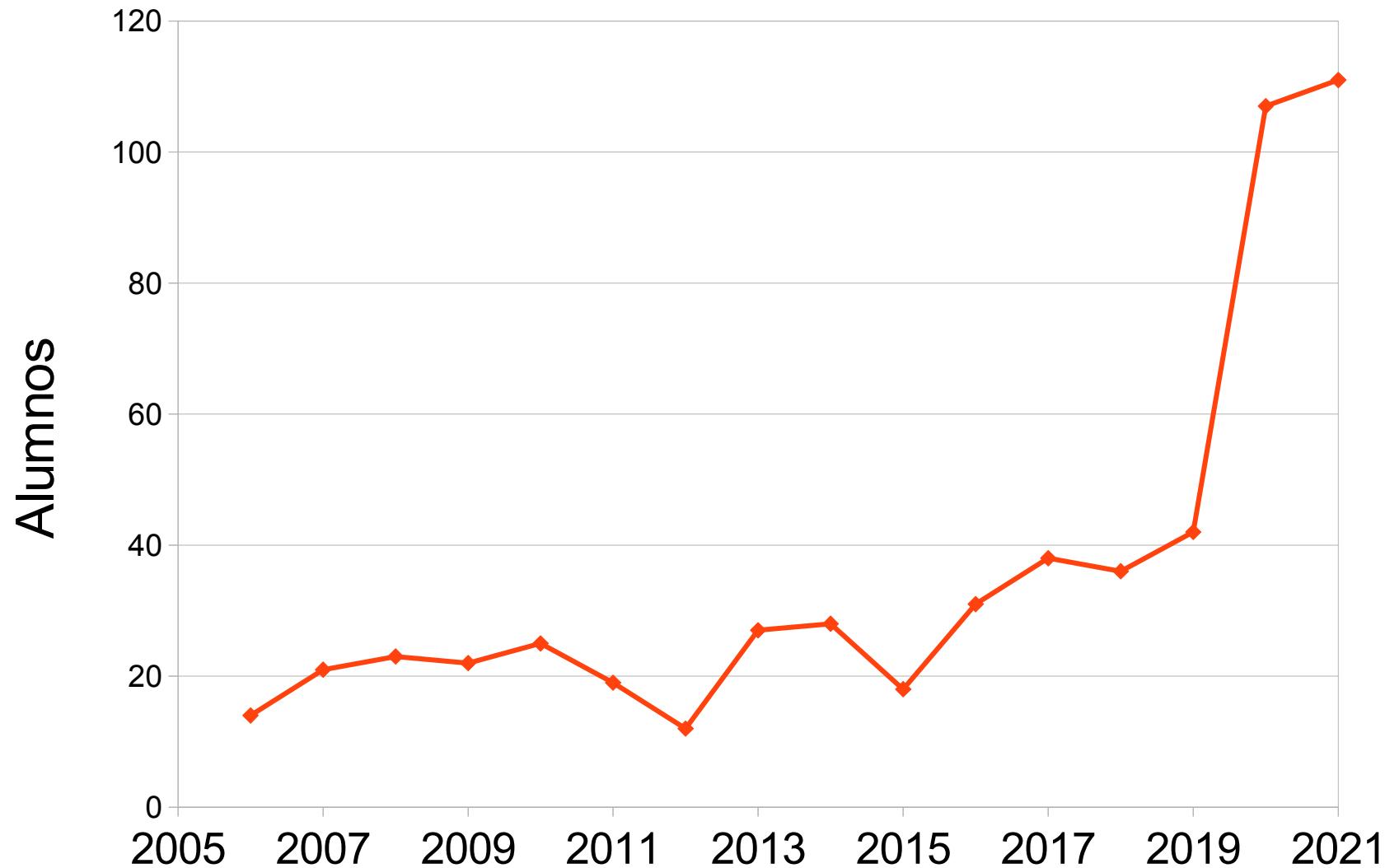
- 1 Despreció la fama para vivir con austeridad y renunció a "El Chavo del 8" por honor: la historia de Ramón Valdés, el actor de "Don Ramón"



- 2 Contundente descargo de la empleada doméstica de Nicole Neumann: "Todo lo que se dijo es mentira"



U. Austral Evolución Cantidad Alumnos



Una carrera de la UBA con trabajo asegurado

Genera gran demanda y buenos sueldos

Nora Bär SEGUIR

29 de enero de 2008



Comentar
(62)



Me gusta



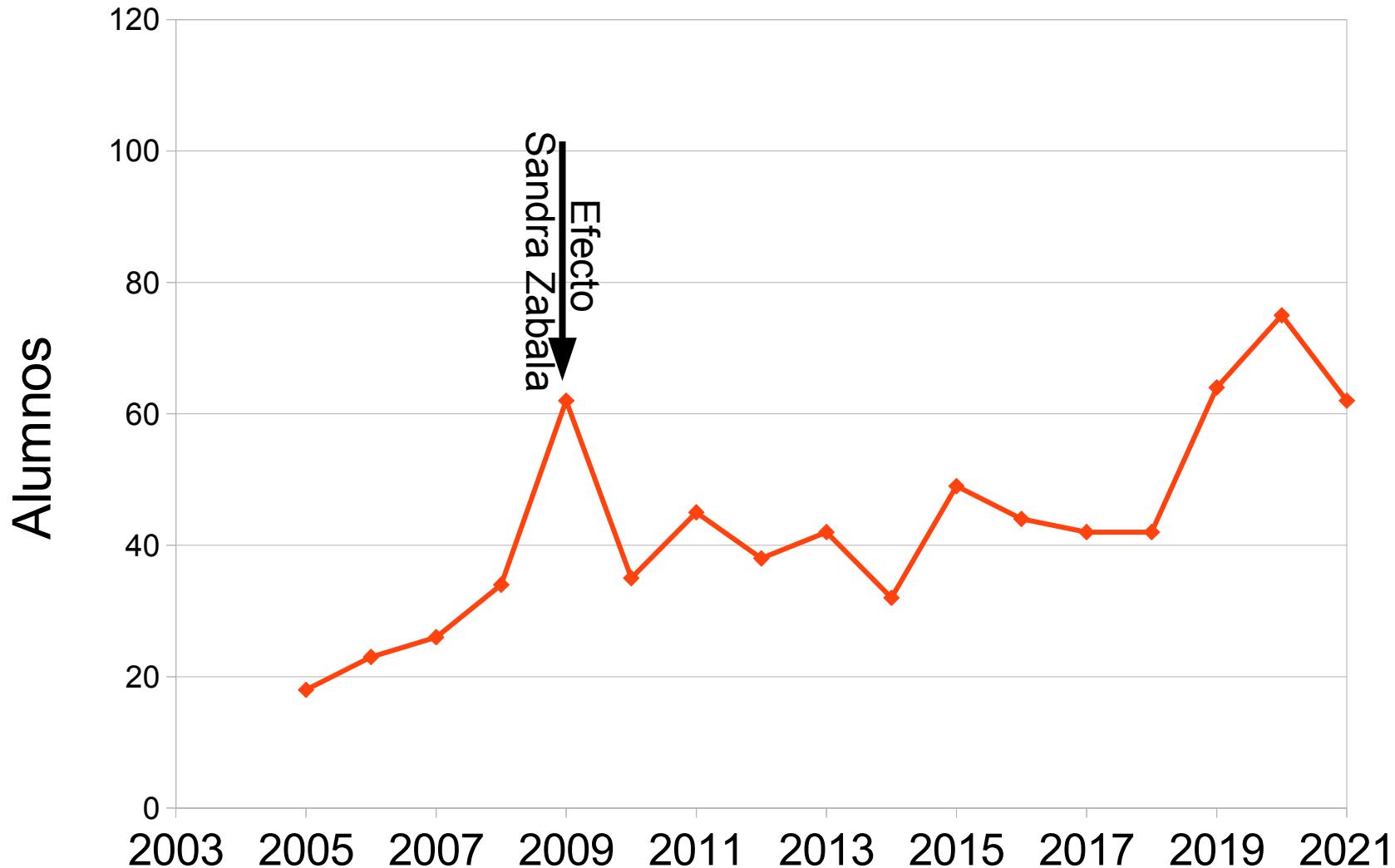
Compartir

Sandra Zabala tiene 37 años, es licenciada en física y trabaja en una agencia de medios. Juan José Lloret tiene 60, es matemático y tiene una consultoría de marketing. Gustavo Markel tiene 30, es ingeniero en sistemas y trabaja en un banco.

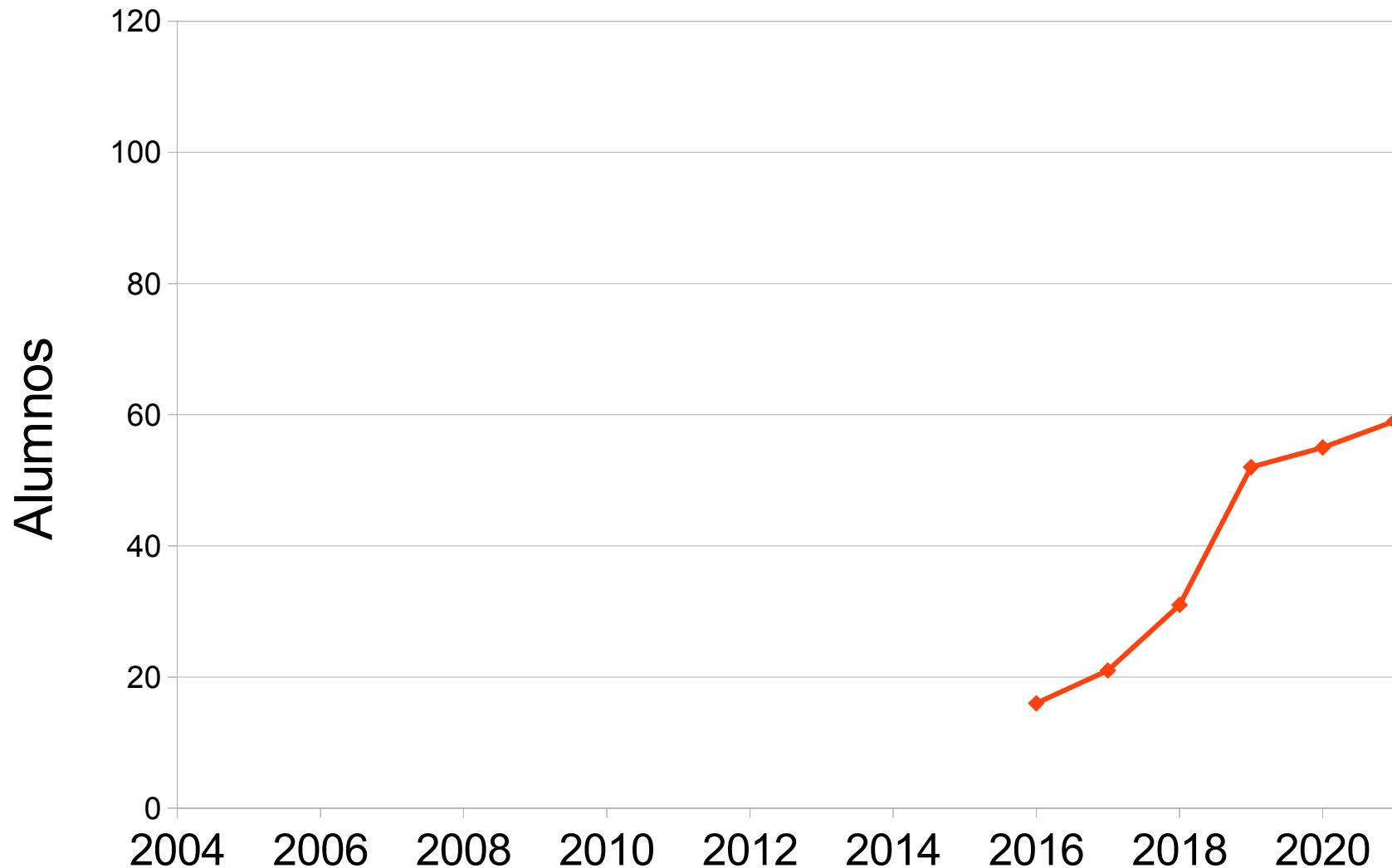
“Estaba trabajando en una empresa, pero el sueldo era paupérrimo (cuenta Sandra Zabala). Ahora soy directora de investigación y desarrollo de una agencia de medios.

En lo económico resultó mejor de lo que había pensado. Multipliqué mi sueldo por 20 o más.”

UBA Evolución Cantidad Alumnos



ITBA Evolución Cantidad Alumnos



¿Usted sabe donde se ha metido?

Maestría Profesional



La maestría profesional se vincula específicamente con el fortalecimiento y consolidación de competencias propias de una profesión o un campo de actuación profesional.

A lo largo del proceso de formación profundiza en competencias en vinculación con **marcos teóricos disciplinares o multidisciplinares** que amplian y cualifican las capacidades de desempeño en un campo de acción profesional o de varias profesiones.

Tipos Maestrías Profesionales

- Especialistas 
- Generalistas

Tipo	Horas Reloj
Diplomatura	80
Especialización	360
Maestría	540
Maestría con Tesis	700
<i>Carrera de Grado</i> S	2600
Carrera de Grado L	3850

Oferta Académica

año 2021

Maestrías y Especializaciones

- Maestría en Análisis y Gestión de Negocios
[Universidad Torcuato di Tella](#)
- Maestría en Ciencia de Datos e Innovación Empresarial, [Universidad CAECE](#)
- Maestría en Explotación de Datos y Gestión del Conocimiento
[Universidad Austral](#)
- Maestría en Ciencia de Datos
[Universidad de San Andres](#)

Maestrías y Especializaciones

- Maestría en Minería de Datos,
[UTN Paraná](#)
- Maestría en Inteligencia de Datos orientada a Big Data, [UNLaPlata](#)
- Especialización en Ciencia de Datos,
[ITBA](#)

Maestrías y Especializaciones

- Maestría en Métodos Cuantitativos para la Gestión y Análisis de Datos en Organizaciones, [UBA Económicas](#)
- Maestría en Explotación de Datos y Descubrimiento del Conocimiento, [UBA Exactas - Ingeniería](#)

Oferta Académica, GRADO

- **ITBA** Licenciatura en Analítica Empresarial y Social , 2019
- **UCEMA** Licenciatura en Analítica de Negocios, 2020
- **UNGB** Licenciatura en Ciencia de Datos, 2020
- **UBA** Licenciatura en Modelado Computacional y Ciencia de Datos, 2021

Heterogeneidad Alumnos

- Expectativas y Objetivos (generalista vs. especialista)
-
- Academia vs Estado vs Industria
- Teoría vs Práctica
- Experiencia técnica
- Edades

Tipo	Horas Reloj
Diplomatura	80
Especialización	360
Maestría	540
Maestría con Tesis	700
Carrera de Grado	3850

El conocimiento necesario para desempeñarse profesionalmente en ciencia de datos yo “no cabe” en las apenas 540 horas de una maestría.

La distancia al objetivo de “saber” depende del punto desde el cual uno parte, con lo cual se puede transitar de muy distinta forma, como el ejemplo que sigue de dos alumnos del año 2019.

36 años, Ing Sistemas, hizo cursos de doctorado,
años de experiencia en R

Laboralmente: desarrollador, luego BI, finalmente
Data Scientist.

En paralelo, docencia.

“Ya por suerte tengo la edad necesaria para
saborear quedarme hasta las 3 am con un buen
café, habano y un LightGBM que no termina de
entender las variaciones de precios del mercado
farmacéutico argentino.

Me llevo un aprendizaje increíble.”

24 años, recien egresado de carrera sociales, jamás programó

“me di cuenta que la maestría no iba a enseñarme lo que yo fui a buscar y tampoco tenían una respuesta clara a: ¿Qué tengo que saber para ser bueno?.

Entonces ahí fue donde descubrí Platzi y fue la manera de mejorar en esto de a poco y llenar nuevamente ese vacío de conocimiento (concreto y aplicable) que la maestría no pudo, ni puede. Con Platzi aprendí Git, Github, Sql, mysql, R, un nivel de python respetable(creo) y ahora recién empiecé con cuestiones referidas al Machine Learning.”

31 años, egresada carrera social

aprender a programar de adulto cuesta más, porque hay que aprender a equivocarse. En otras disciplinas no hay carteles de error, por ejemplo un sociólogo puede escribir el paper mas idiota del mundo sin que nadie se lo diga, en cambio si tu código esta mal, esta mal, no compila, no corre.

Parece una bobada pero en mi experiencia es el elemento más perturbador para gente proveniente de carreras humanísticas o simplemente no exactas.

32 años, ingeniero (no IT)

soy ingeniero y mi trabajo es técnico, usualmente programo y estoy acostumbrado a analizar y hacer experimentos.

Sin embargo **siento que esa misma formación me jugó en contra**, ya que en mi disciplina experimentamos con cosas que creemos que van a funcionar, que la teoría sugiere que funcionan, no probamos cualquier cosa sin sentido.

Por eso me limité en las pruebas y otros compañeros más arriesgados obtuvieron mucho mejores resultados. Aún no me imagino como empezar a pensar out of the box.

La Asignatura

Se presentará un caso de estudio con dataset de clientes de empresa financiera .

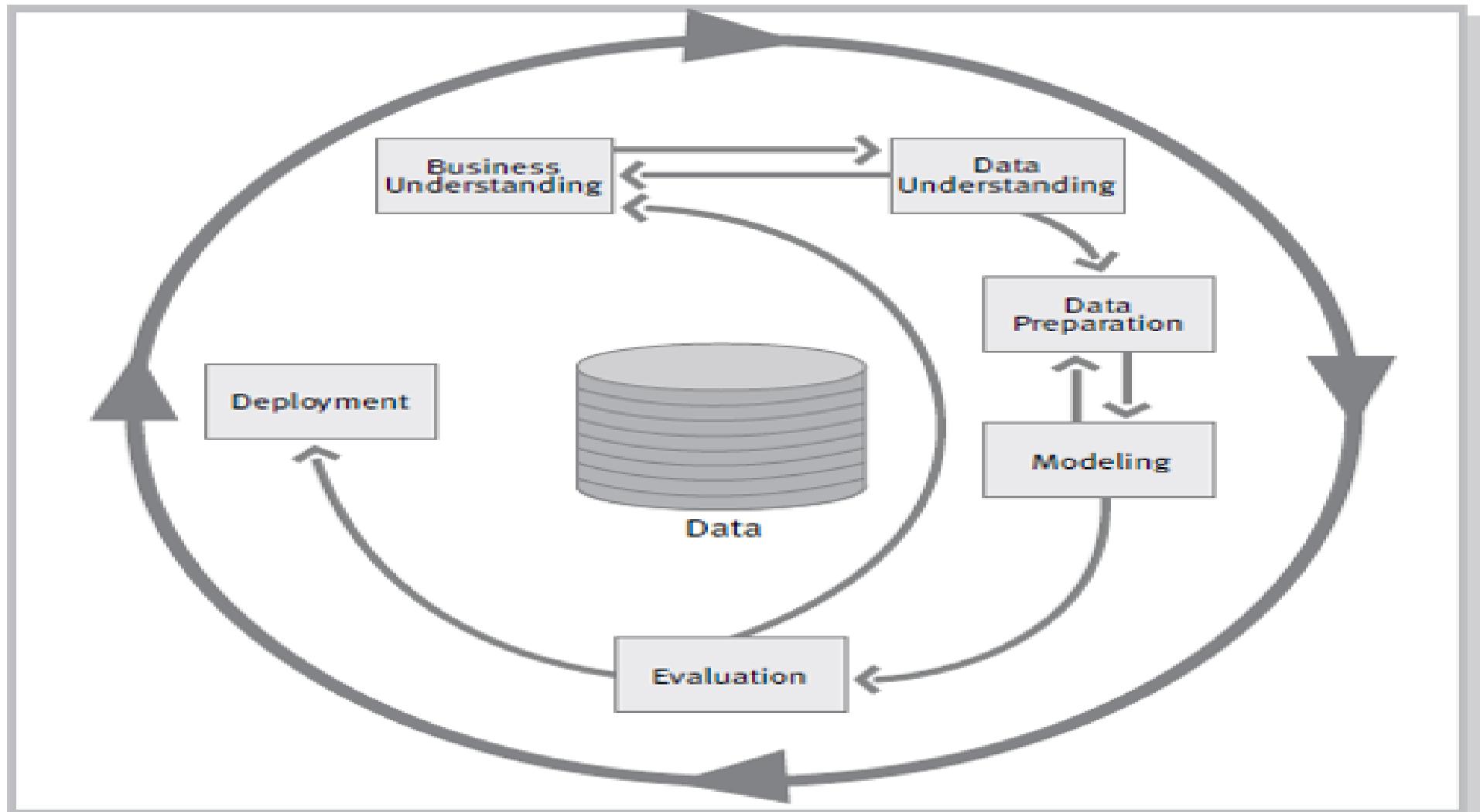
Se guiará a los alumnos a desarrollar un modelo predictivo de attrition que sea competitivo con los de nivel profesional existentes en el mercado argentino.

Se organiza alrededor de DOS competencias.

Data Science Tasks

- Description
- Prediction
 - Structured Data 
 - Unstructured Data
- Causal Inference

Nuestra GRAN limitación



Modalidad de Dictado

Se hará el planteo del problema entregando el dataset y comentando en gran detalle su historia y las motivaciones, precisando claramente el objetivo a resolver.

Luego se acompañará a los alumnos en la resolución, presentando los elementos teóricos y conceptuales en el justo tiempo que van siendo requeridos para solucionar las distintas partes del problema.

Manifiesto Pedagógico

Enseñar,
más que llenar un recipiente
es **encender** un fuego.

Plutarco

- aprendizaje basado en pares
 - observar y aprender de sus pares
-
- recorrer su propio camino
 - diseñar creativamente sus propios experimentos



kaggle



GitHub



git

The problem is not people being uneducated.

The problem is that people are educated just enough
to believe what they have been taught,
but not educated enough to **question anything from**
what they have been taught.

Be undisciplined. Be original. Be irreverent.

Richard Feynman

Estoy aquí para empujar a las personas
más allá de lo que se espera de ellas.

How, when and why students learn meaningfully, or just regurgitate facts and deploy procedures and algorithms (or possibly don't manage even those). We only think deeply about things we care about.

Emotion is the rudder of thought.

Meaningful learning is actually about helping students to connect their isolated algorithmic skills to abstract, *intrinsically emotional*, subjective and meaningful experiences.

Mary Helen Immordino-Yang

Free Speech Statement, April 2018

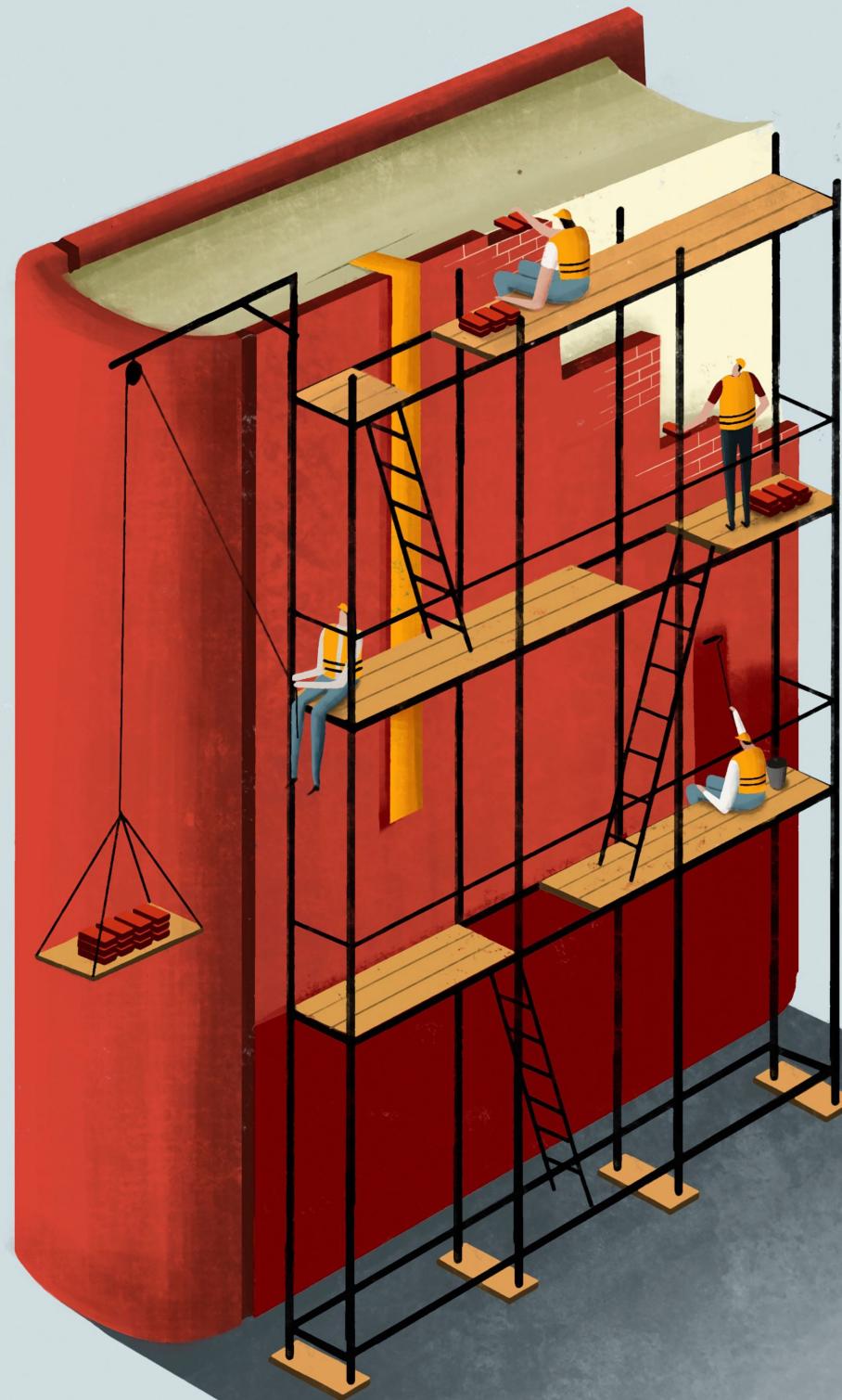
The free and open exchange of ideas and information is fundamental to the educational mission of *Association of American Universities*. The robust discussions and debates that occur at research universities have been central to the advancement of democracy, the creation of new knowledge, the fostering of educational excellence, and the promotion of social progress. As heads of these institutions we are unequivocally committed to preserving and honoring this proud heritage.

While we may deem some speech to be odious, disgraceful, and antithetical to our values, our campuses are and should remain places where ideas can be expressed free of disruption, intimidation, and violence.

Few academics challenge censorship that emerges from students. It is important that more do, because a culture that restricts the free exchange of ideas encourages self-censorship and leaves people afraid to express their views in case they may be misinterpreted. This risks destroying the very fabric of democracy.

An open and democratic society requires people to have the courage to argue against ideas they disagree with or even find offensive. At the moment there is a real risk that students are not given opportunities to engage in such debate.

A generation of students is being denied the opportunity to test their opinions against the views of those they don't agree with.



FIN