



Machine Learning :

De 0 a Hero

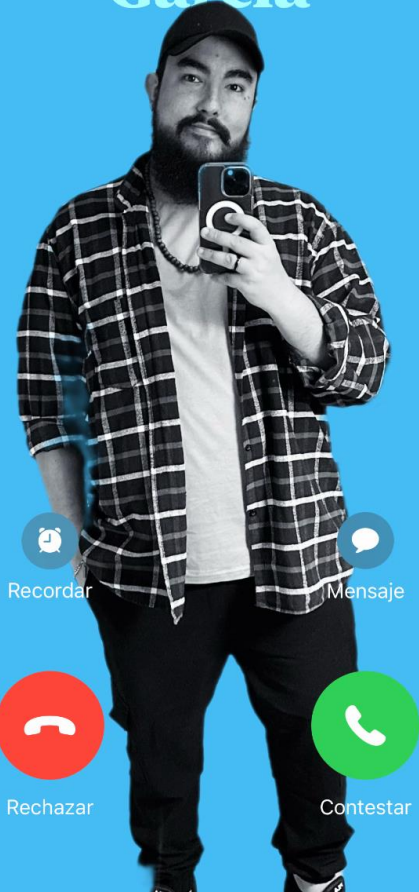
Alvaro Garcia

TL Data at  CloudHesive

- 13 Años en consultoría estratégica de datos.
- Entusiasta de la seguridad informática.
- Uruguayo.



Alvaro
Garcia



Recordar

Mensaje

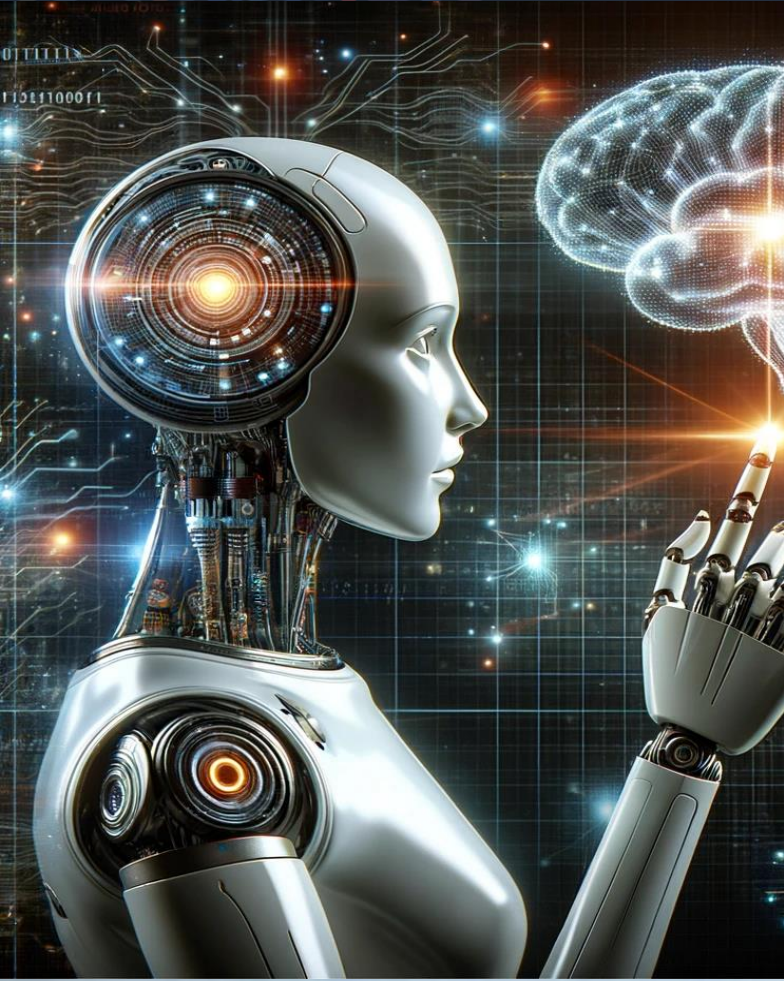
Rechazar

Contestar

AGENDA

- Que es la Inteligencia Artificial ?
- Comprendiendo los conceptos clave
- Skills
- Fases de un proceso ML
- Herramientas AWS
- Conclusiones
- Q&A

Que es la Inteligencia Artificial ?



Inteligencia Artificial

- Es el campo de estudio que permite a las máquinas imitar las capacidades cognitivas humanas.
- Es la creación de sistemas que pueden aprender y actuar de forma autónoma.
- Algoritmos que simulan inteligencia para resolver problemas

Un poco de historia

50' a 60'

- Prueba de Turing
- Conferencia de Dartmouth.
- Primeros programas de IA.

70' a 80'

- Primer Invierno de la IA.
- Sistemas expertos.
- Segundo Invierno de la IA.

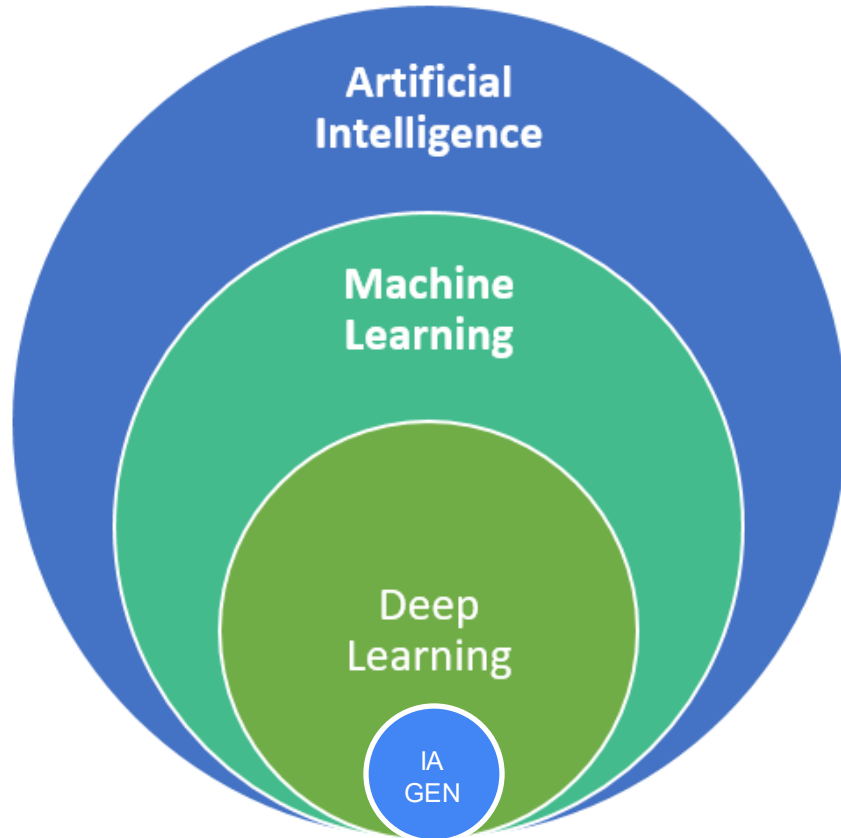
90' a 00'

- Mejoras de algoritmos.
- Big Data
- Simplificación del ML
- AWS

10' a 20'

- Deep Learning
- Watson & AlphaGo
- IA generativa
- Ética y Regulación

$ML \subseteq IA$ (ML está contenido dentro de IA)

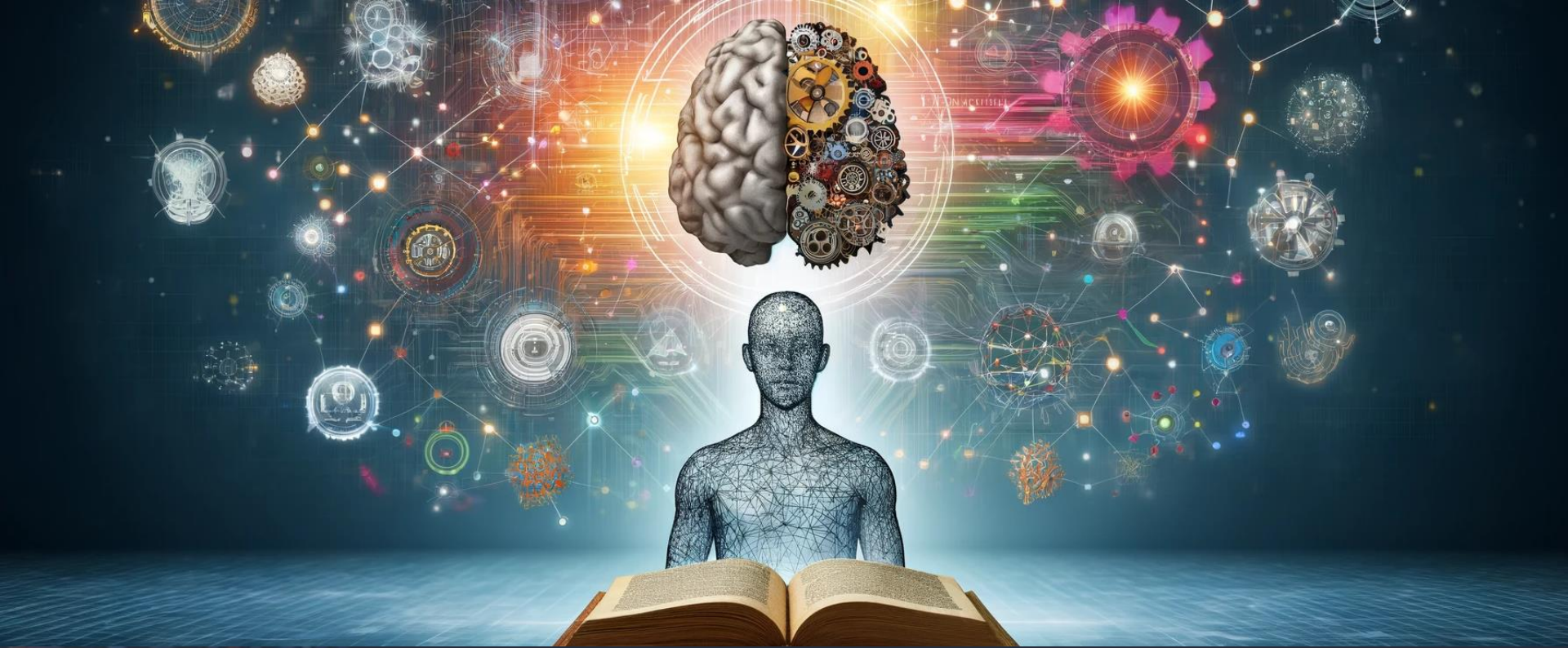


Algoritmos que simulan al humano

Algoritmos que aprenden de datos acotados

Simulan una red neuronal y trabajan con gran cantidad de datos

Son capaces de crear a partir de sus datos de entrenamiento



Comprendiendo los conceptos clave

A que se dedica un data scientist ?



No es solo desarrollar modelos de ML

No es solo hacer visualizaciones

No es solo escribir código

**Generar impacto en el
negocio utilizando
datos**

Impacto en el negocio



Insights 

Data Products 

Nuevos
productos/servicios 



Skills

De un Data Scientist

Matemática

Cálculo

- Derivadas para optimización algorítmica.
- Integrales evalúan acumulaciones.
- Fundamental en machine learning.

Álgebra

- Manejo de matrices, vectores.
- Clave en sistemas recomendación.
- Soporta operaciones a gran escala.

Estadística

- Resumen visual de datos.
- Fundamento de modelos predictivos.
- Impulsa decisiones basadas en datos.

Probabilidad

- Modela incertidumbre y predicción.
- Base para clasificaciones estadísticas.
- Esencial para inferencia bayesiana.

Python

- Principal lenguaje para Data science
- Gran comunidad y documentación
- Muchas herramientas disponibles

SQL

- Saber navegar por una base de datos
- Unificación y Normalización
- Optimización de tiempos

Git

- Control de versiones.
- Facilita la colaboración en proyectos.
- No integra el código y la documentación.

Jupyter Notebook

- Escribir y ejecutar código.
- Documentar el análisis.
- Cambiar y visualizar resultados en vivo.

Procesos

- Estructura organizacional
- Cultura
- Procesos internos y externos
- Time to market
- Retorno de la inversión

Productos/Servicios

- Que se vende ?
- Como se vende ?
- Quien es el cliente ?
- Competidores

**Saber a que se dedica y
como trabaja el
negocio**



Fases de un proceso ML



Recolección

Y Almacenamiento

Fase 1 - Recolección & Exploración

Detectar

Identificar las fuentes de datos, obtener su ubicación, velocidad de creación, volumen, etc.



Procesar

Realizaremos todos los ETL necesarios hasta obtener el dato en el formato y estructura necesario.



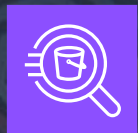
Analizar

Con los datos modelados analizaremos los datos en búsqueda de sesgos y/o patrones e identificaremos las variables importantes.

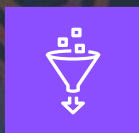


Fase 1 - Recolección & Exploración

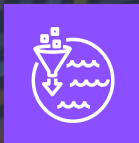
Detectar



Amazon Athena



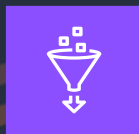
AWS Glue



AWS Lake Formation



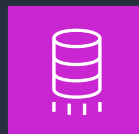
Procesar



AWS Glue



Kinesis



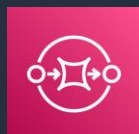
AWS DMS



AWS Lambda



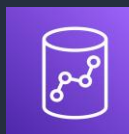
EMR



Amazon SQS



S3



Redshift

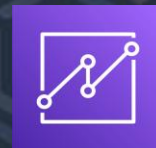


AWS CDK

Analizar



Amazon SageMaker



Amazon QuickSight



Exploración

Y Agregación de datos

Fase 2 - Exploración y agregación

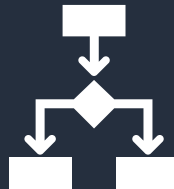
Limpieza

Aplicaremos reglas a los datos que estén mal cargados, con errores para que se pueda representar lo más posible la realidad.



Preparación

Este proceso implica organizar, limpiar y convertir los datos en formatos adecuados para facilitar un entrenamiento precisos y eficiente.



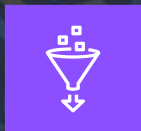
Anomalías

Identificar datos atípicos que revelan errores o insights importantes.

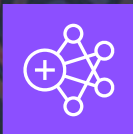


Fase 1 - Recolección & Exploración

Limpieza



AWS Glue



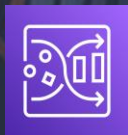
EMR



AWS Lambda

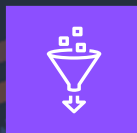


Sage Maker



Data Brew

Preparación



AWS Glue



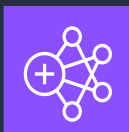
Kinesis



Sage Maker



AWS Lambda



EMR

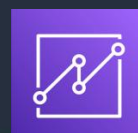


Data Pipelines

Anomalías



Amazon SageMaker



QuickSight



Amazon Macie



Kinesis DA



Aprendizaje

Y exploración

Fase 3 - Aprendizaje y Exploración

Machine Learning

En base a nuestro caso de uso seleccionaremos uno o varios modelos de ML que permitan resolver el problema que tenemos.



A/B Testing

A la hora de tener disponible nuestro modelos realizaremos pruebas reales que permitan confirmar la utilidad del modelo.



Evaluación

De manera iterativa y constante se debe trabajar con los modelos en pos de ajustarlos con el tiempo.



Fase 3 - Aprendizaje y Exploración

ML



Amazon SageMaker



Notebook



Canvas



Model



Train

A/B Testing



Amazon SageMaker



AWS Lambda

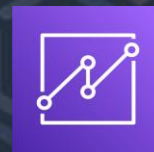


API GW

Evaluación



Amazon SageMaker



Amazon QuickSight



Deep learning

e IA Generativa

Fase 4 – Deep learning e IA Generativa

Deep Learning

Cuando nuestro caso de uso involucra algo más que datos (imágenes, video, etc), debemos utilizar modelos de aprendizaje profundo en nuestra estrategia para la solución.

IA Generativa

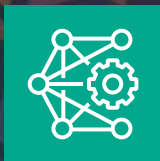
Cuando exista la necesidad de generar respuestas en base a datos existentes, interactuar con clientes, interpretar consultas o modelar datos para acelerar un proceso.

Fase 4 - Aprendizaje y Exploración

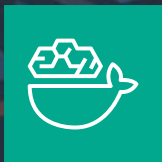
Deep Learning



Amazon SageMaker



AWS Deep Learning AMIs



AWS Deep Learning
Containers



TensorFlow on AWS

IA Generativa



Amazon Bedrock



Amazon Q



Resumen de servicios IA en AWS

Resumen de servicios IA en AWS

AI SERVICES



NEW

Amazon
HealthLake

HEALTH AI



Amazon
Transcribe
Medical



Amazon
Comprehend
Medical



NEW

AWS Panorama
+ Appliance



NEW

Amazon
Monitron

INDUSTRIAL AI



NEW

Amazon Lookout
for Equipment



NEW

Amazon Lookout
for Vision

ANOMALY DETECTION



NEW

Amazon Lookout
for Metrics



NEW

Amazon
DevOps Guru



Amazon
CodeGuru

VISION



Amazon
Rekognition

SPEECH



Amazon
Polly



Amazon
Transcribe

+Medical

TEXT



Amazon
Comprehend

+Medical



Amazon
Translate



Amazon
Textract

SEARCH



Amazon
Kendra

CHATBOTS



Amazon
Lex

PERSONALIZATION



Amazon
Personalize

FORECASTING



Amazon
Forecast

FRAUD



Amazon
Fraud Detector

CONTACT CENTERS



Contact Lens

Voice ID

For Amazon Connect

ML SERVICES



Amazon
SageMaker

Label
data

NEW

Aggregate &
prepare data

NEW

Store & share
features

Auto ML

Spark/R

NEW

Detect
bias

Visualize in
notebooks

Pick
algorithm

Train
models

Tune
parameters

NEW

Debug &
profile

Deploy in
production

Manage
& monitor

NEW

CI/CD

Human review

SAGEMAKER STUDIO IDE

NEW: SageMaker JumpStart

NEW: Model management for edge devices

FRAMEWORKS & INFRASTRUCTURE



TensorFlow



mxnet



PyTorch



GLUON



Keras



learn



DeepGraphLibrary

Deep
Learning
AMIs &
Containers

GPUs &
CPUs

Elastic
Inference

Trainium

Inferentia

FPGA



Ética

Ética en el manejo de datos



Do

Consentimiento claro antes de recopilar/utilizar cualquier dato personal.

Conocer la normativa vigente. (RGPD, Argentina Ley 25326, Chile Ley 19628)

Protección rigurosa contra el acceso no autorizado.

Dont's

Uso de datos sin consentimiento explícito del usuario.

Vender/Utilizar información personal sin permiso del titular.

Ignorar solicitudes de eliminación de datos personales.

**Nunca trabajar con datos de
los que se desconoce el origen**



Data Engineer



Data Analysis



Charts



Machine Learning



Machine Learning



Tareas y roles

Tareas y roles

APRENDER Y OPTIMIZAR

AGREGAR Y ETIQUETAR

EXPLORAR Y TRANSFORMAR

MOVER Y ALMACENAR

RECOLECTAR



Tareas y roles – Start UP

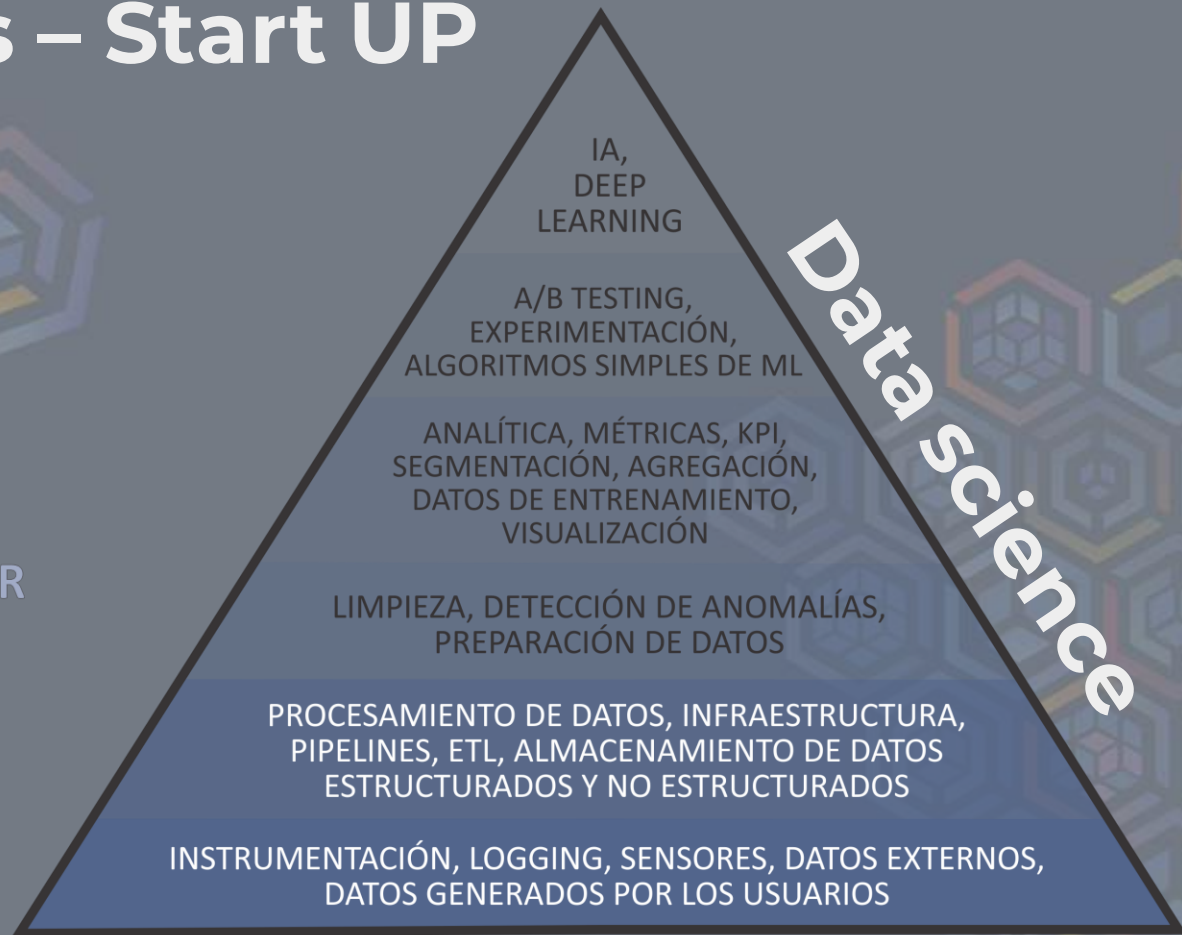
APRENDER Y OPTIMIZAR

AGREGAR Y ETIQUETAR

EXPLORAR Y TRANSFORMAR

MOVER Y ALMACENAR

RECOLECTAR



Tareas y roles - PYME

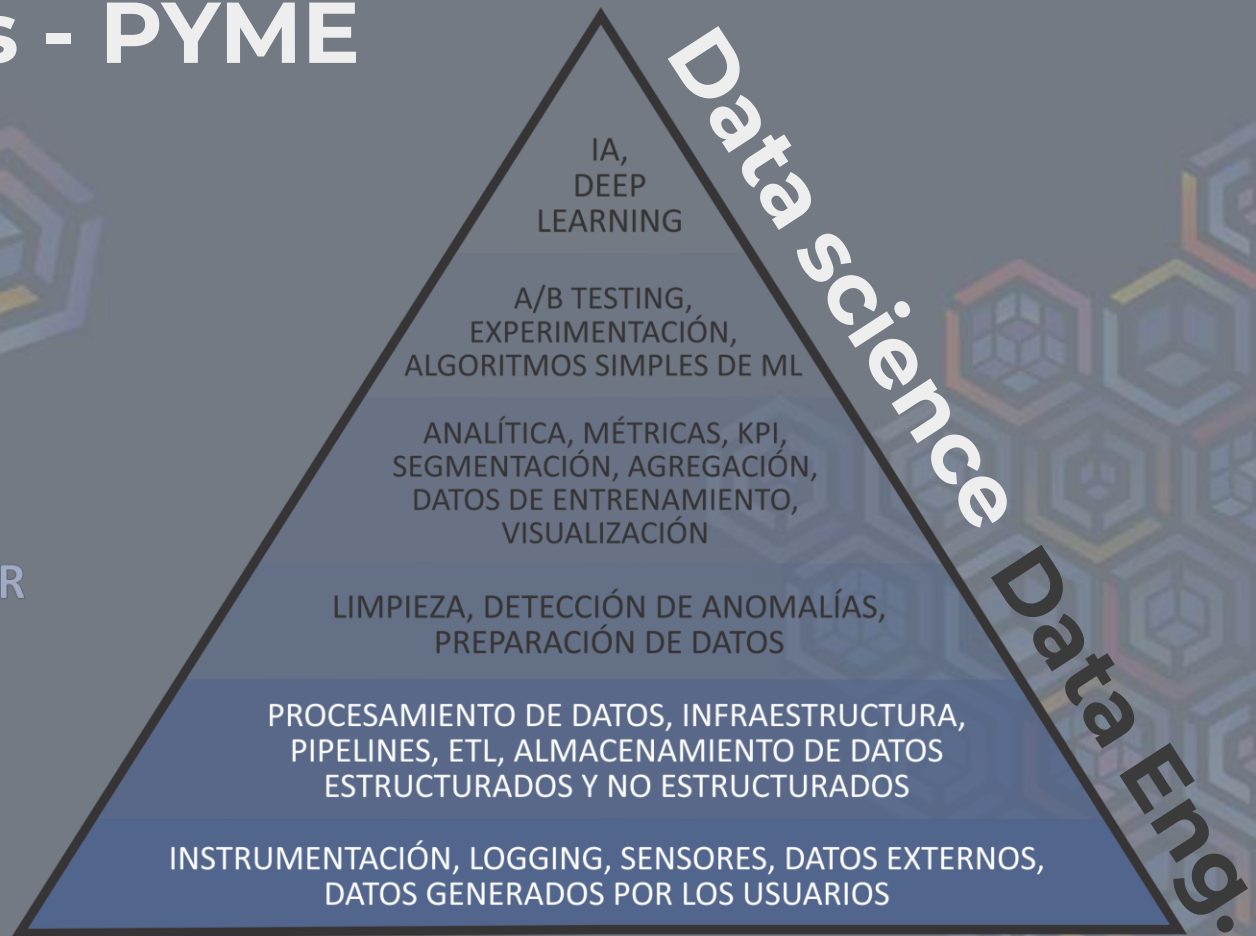
APRENDER Y OPTIMIZAR

AGREGAR Y ETIQUETAR

EXPLORAR Y TRANSFORMAR

MOVER Y ALMACENAR

RECOLECTAR



Tareas y roles – Grande

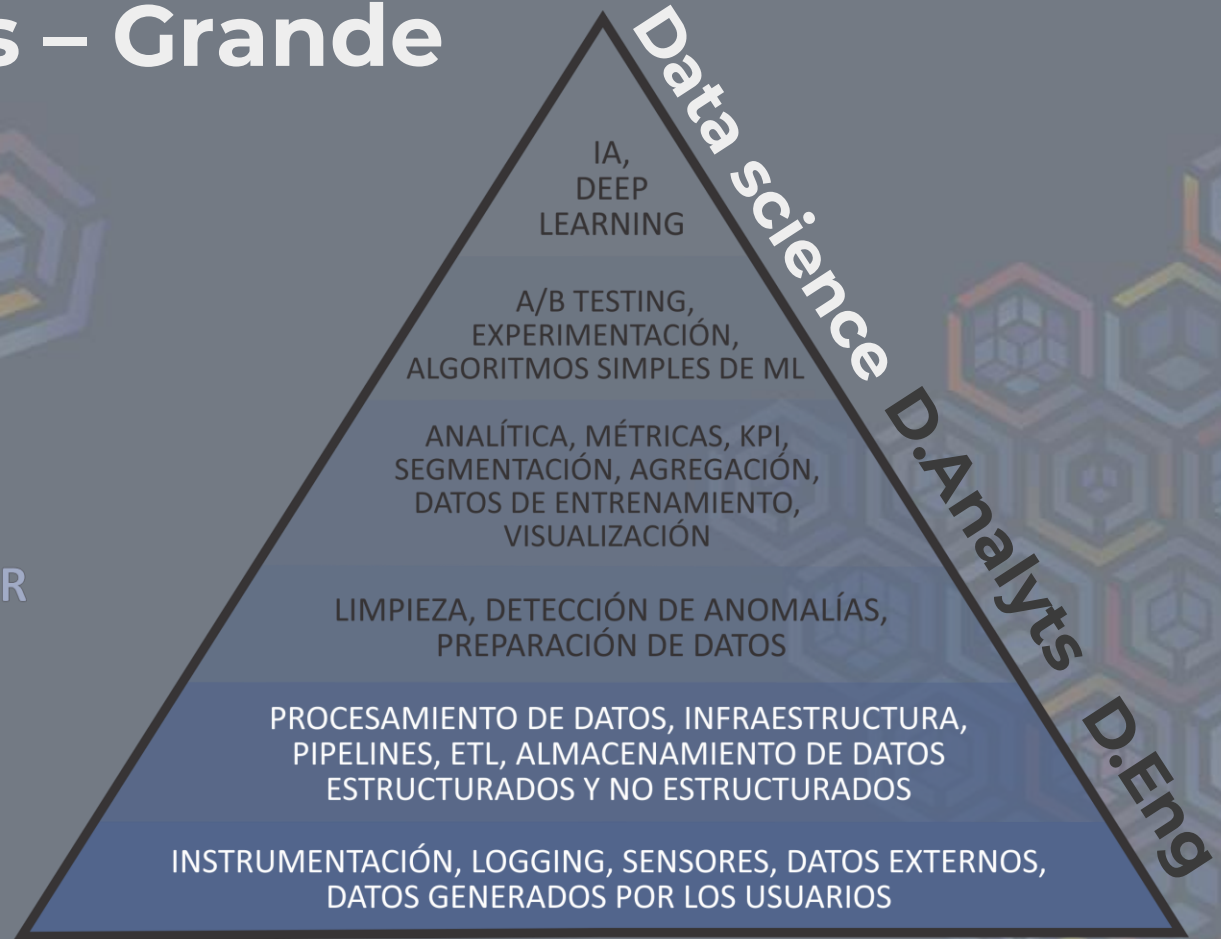
APRENDER Y OPTIMIZAR

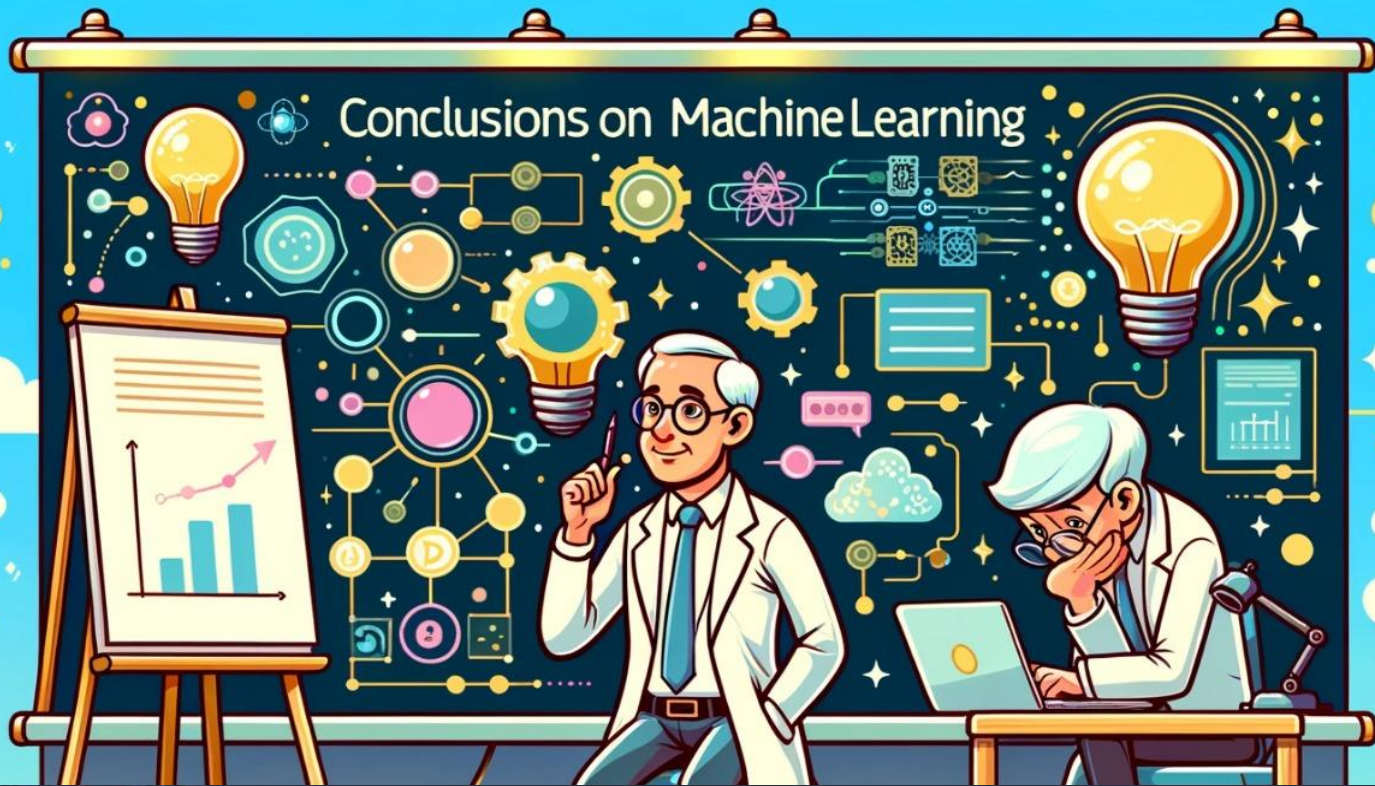
AGREGAR Y ETIQUETAR

EXPLORAR Y TRANSFORMAR

MOVER Y ALMACENAR

RECOLECTAR





Conclusiones

Conclusiones

- Tareas ambiguas 🤔
- Modelo entrenado es modelo viejo 🕒
- Experimentar (POC / MVP) 🧪
- No utilizar datos que desconozcamos 🚫 📊
- Investigar antes de desarrollar 🔍 📚



**NO HAY QUE SER UN EXPERTO PARA
SER UN GRAN** Data Scientist

¡Muchas gracias!

SE INICIA LA RONDA DE PREGUNTAS.

