

# Ciência de dados na prática com Oracle Data Science

Erik Gama

Trilha Inovação com dados em nuvem  
03.12.2020 17h10

# Ciência de dados na prática com Oracle Data Science

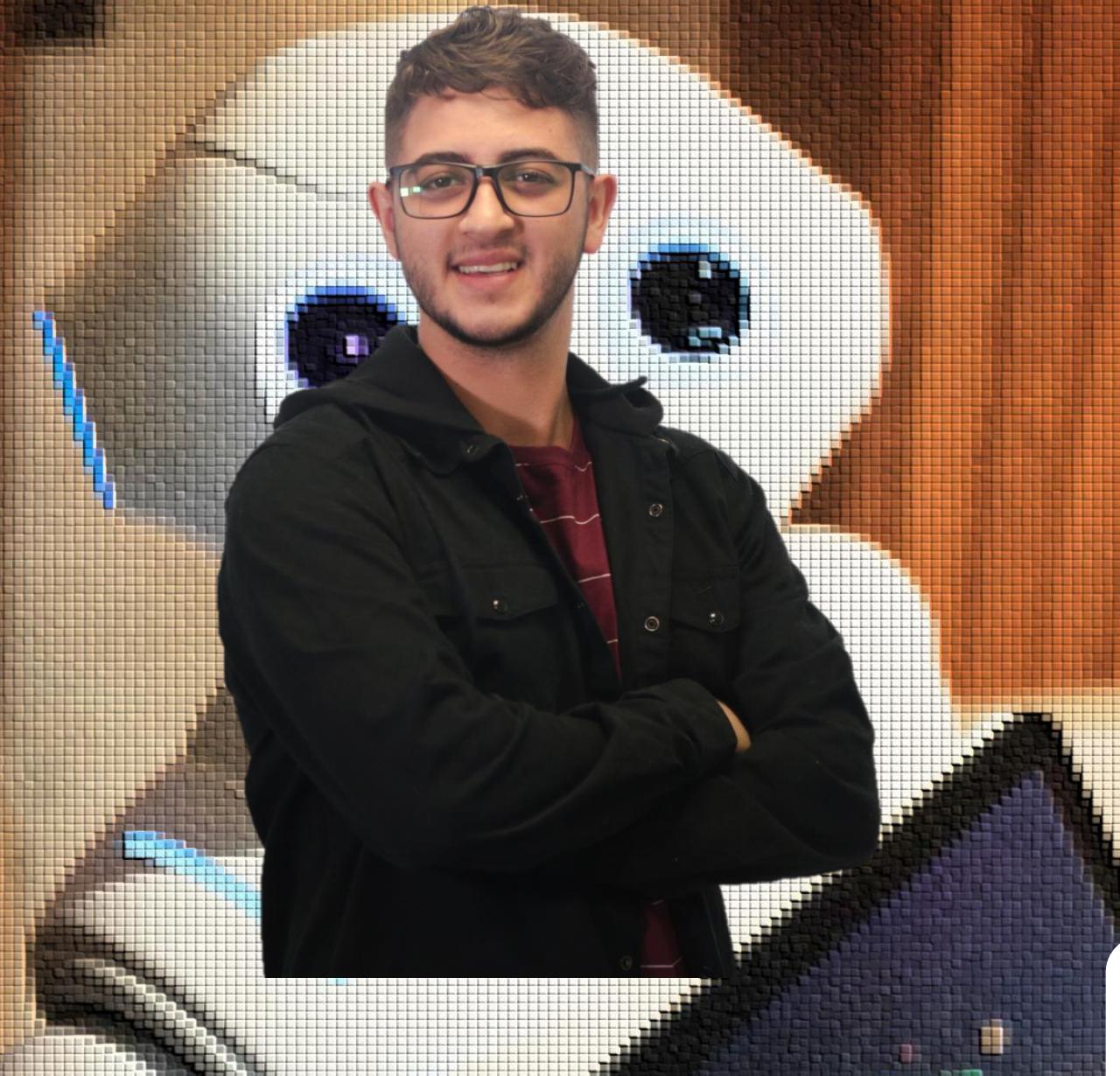
Erik Gama

Trilha Inovação com dados em nuvem  
03.12.2020 17h10



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhável 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.

# Erik Gama



### Origens de Dados



### Integração



### Persistência dos Dados



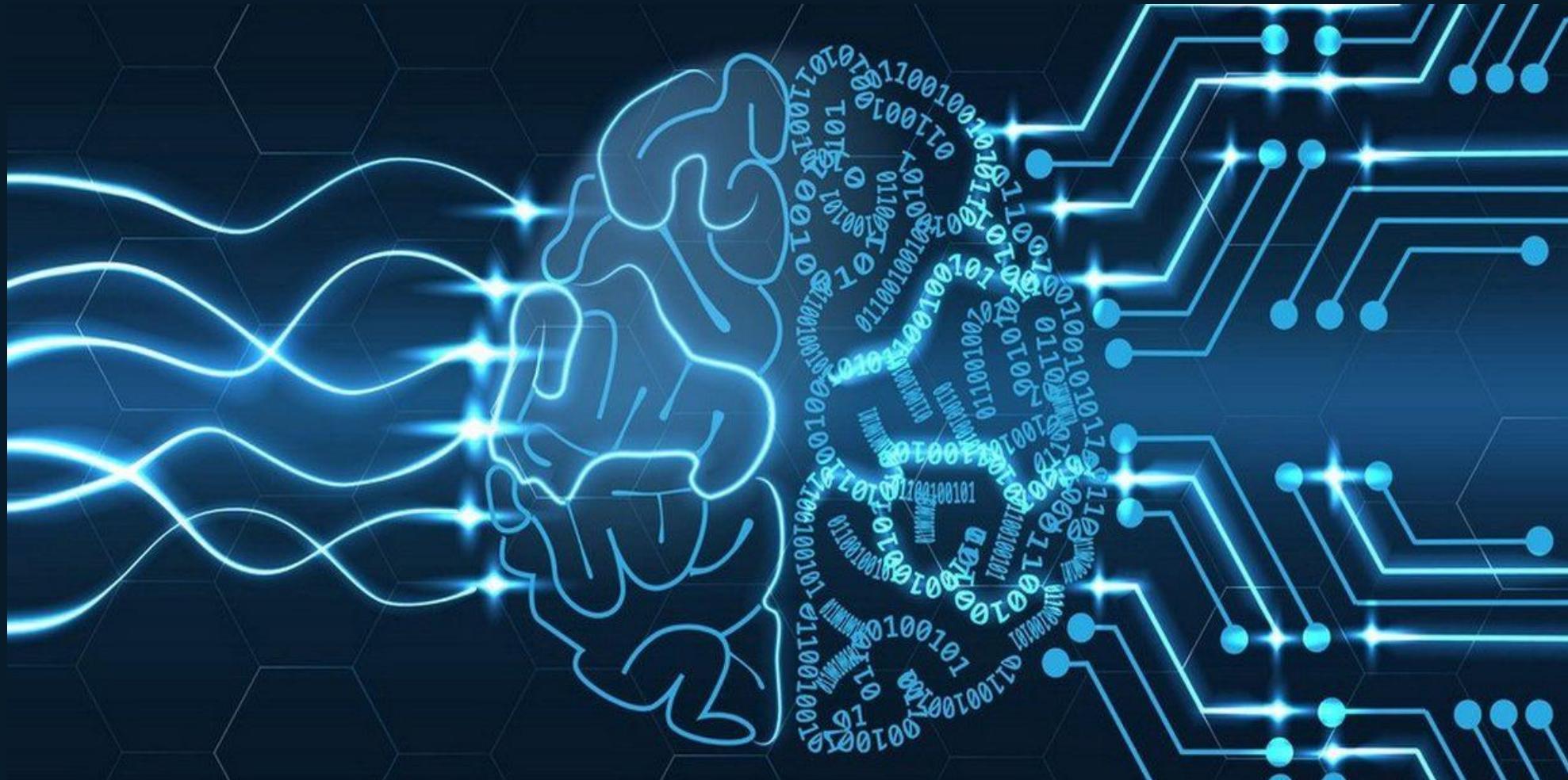
Data Warehouse

### Analise de Dados



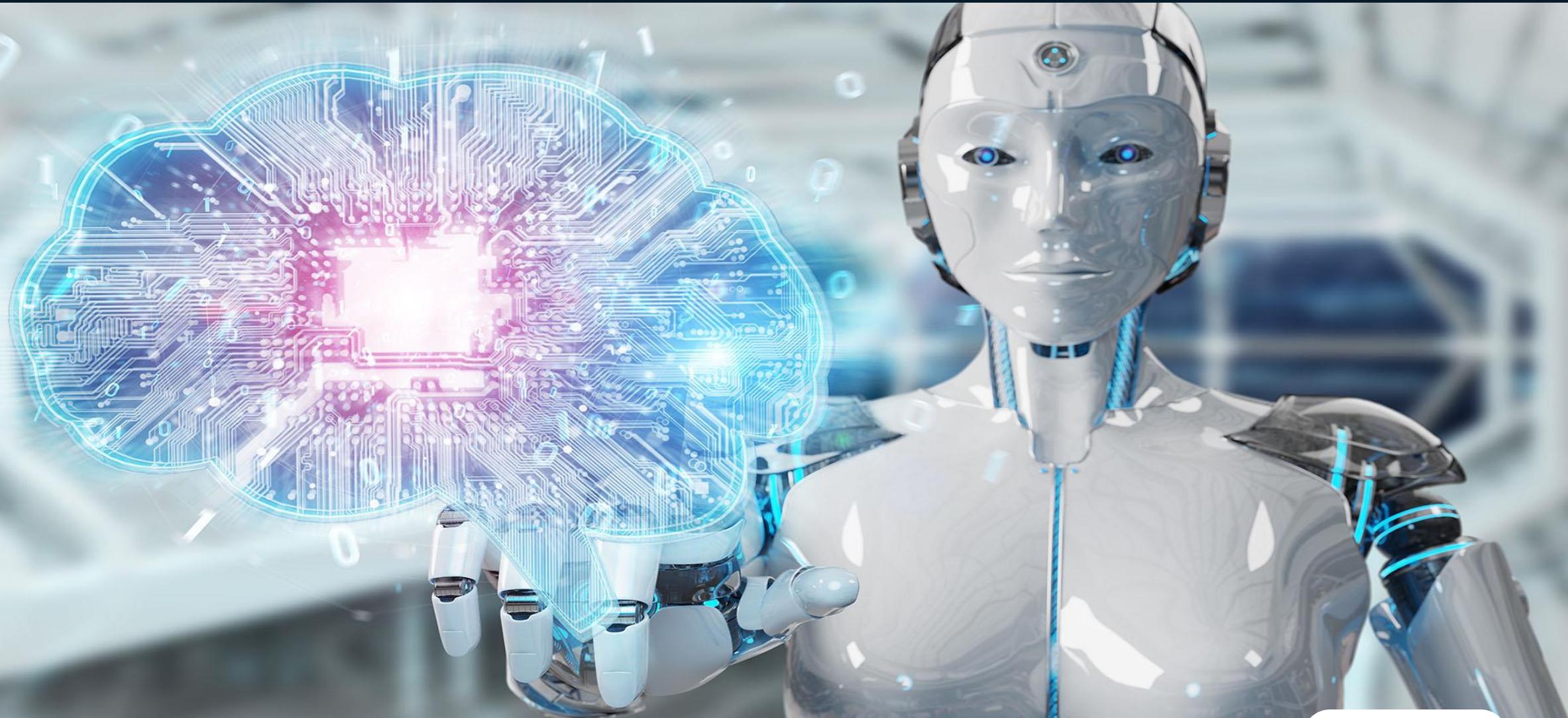


A.I.



**PROBLEM** →  → **SOLUTION**  
**IA**

01001001 00101110 01000001 - 01001001 00101110 01000001 - 01001001 00101110 01000001

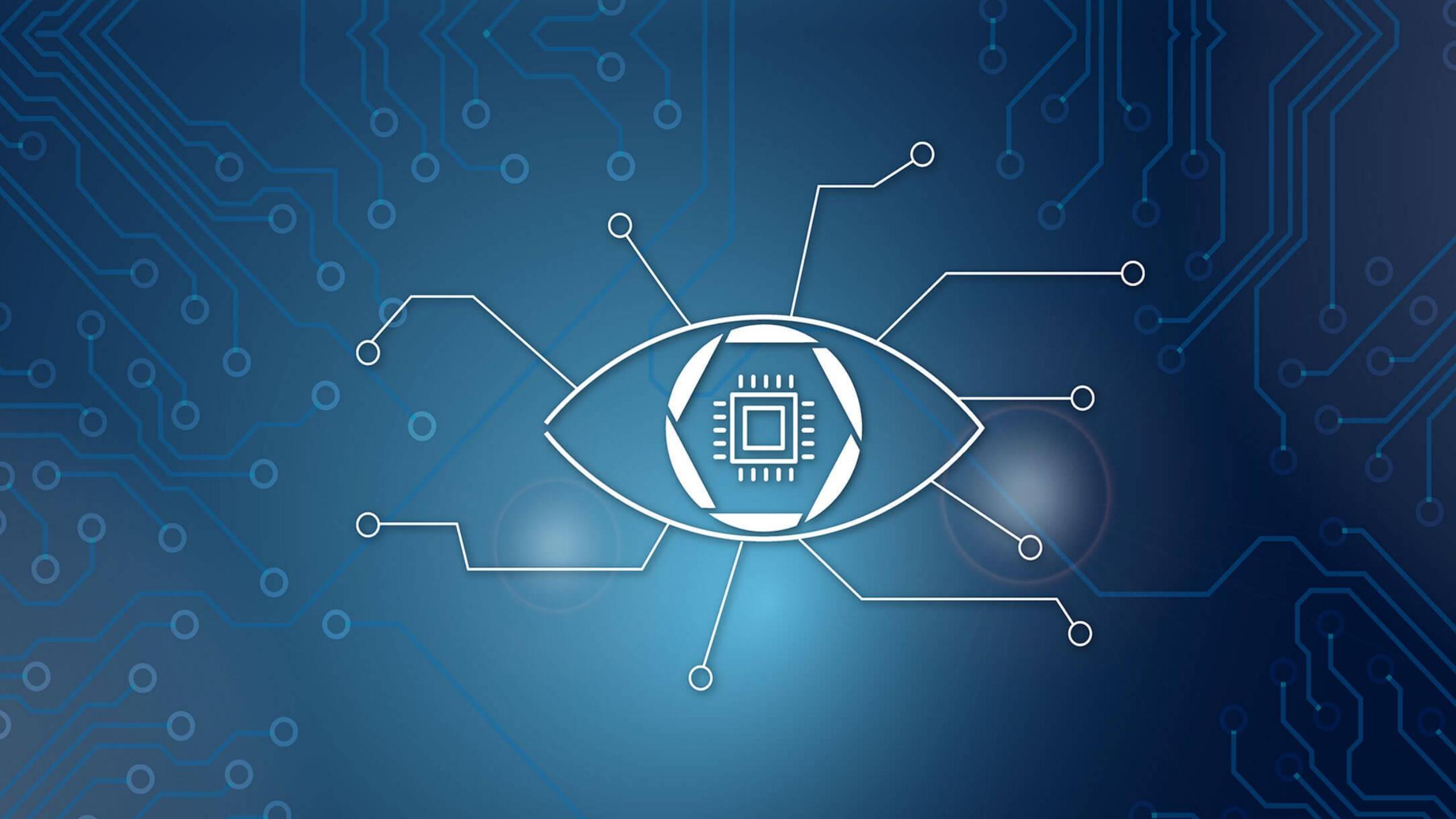


01001001 00101110 01000001 - 01001001 00101110 01000001 - 01001001 00101110 01000001



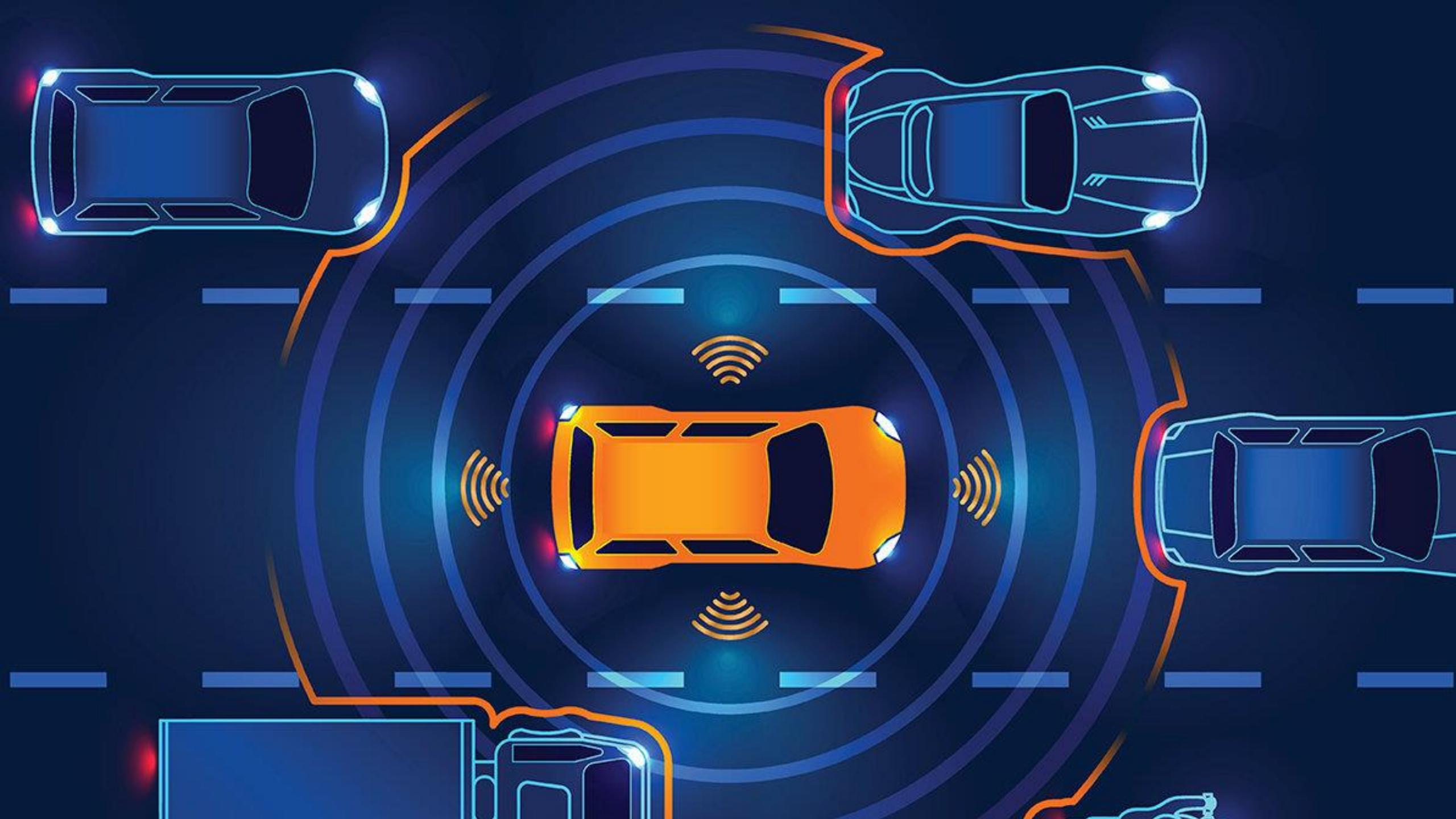
# ONDE as I.A. são APLICADAS?

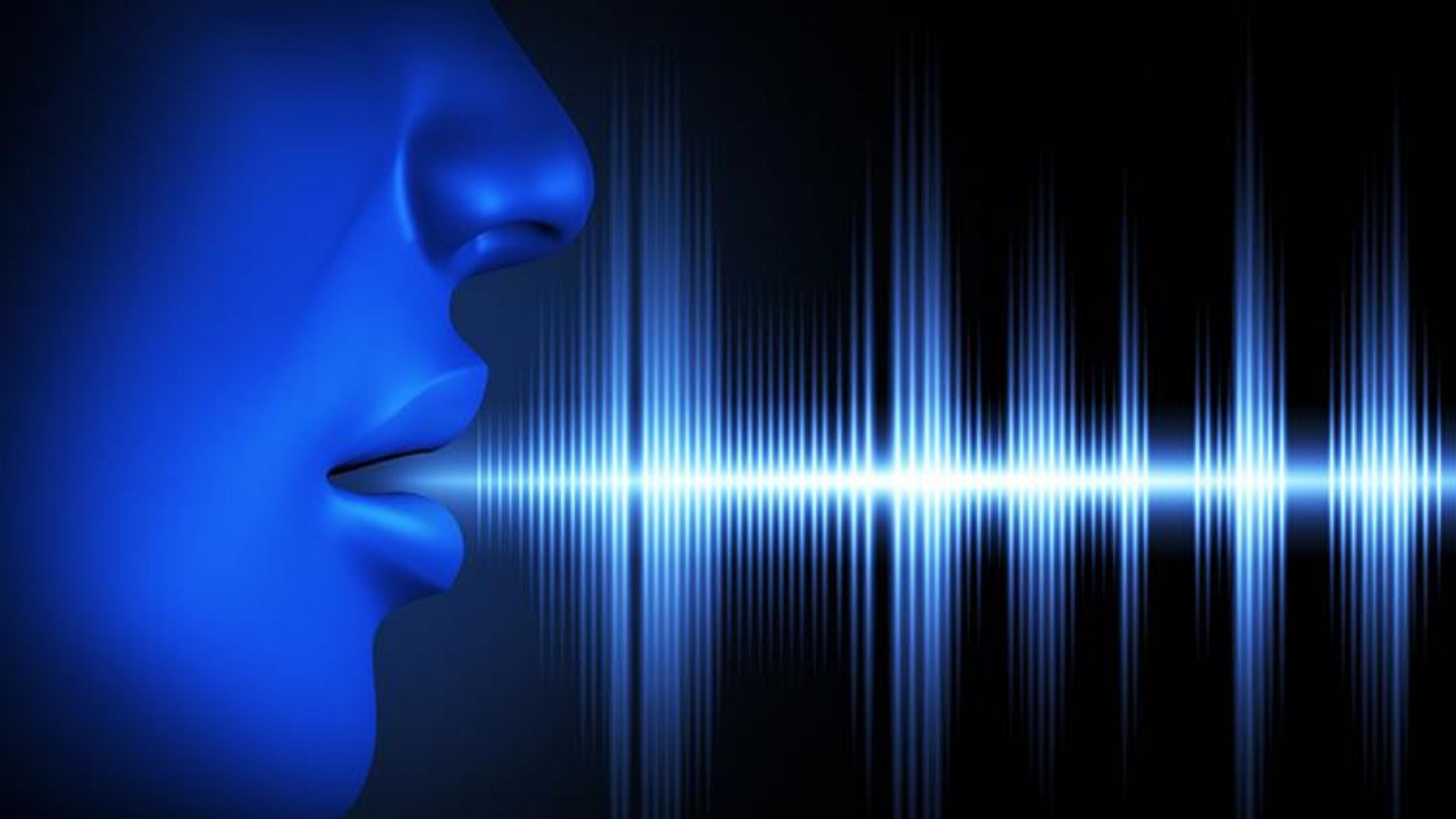


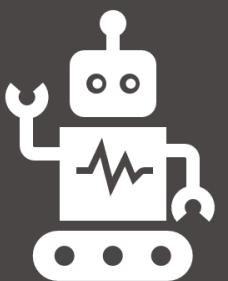
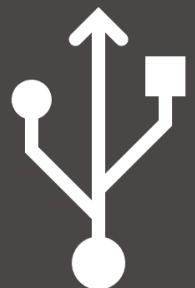
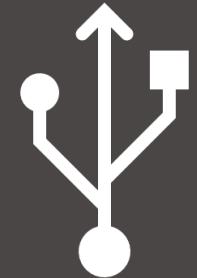
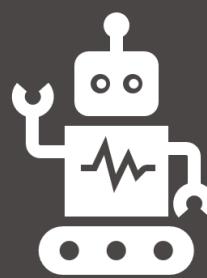








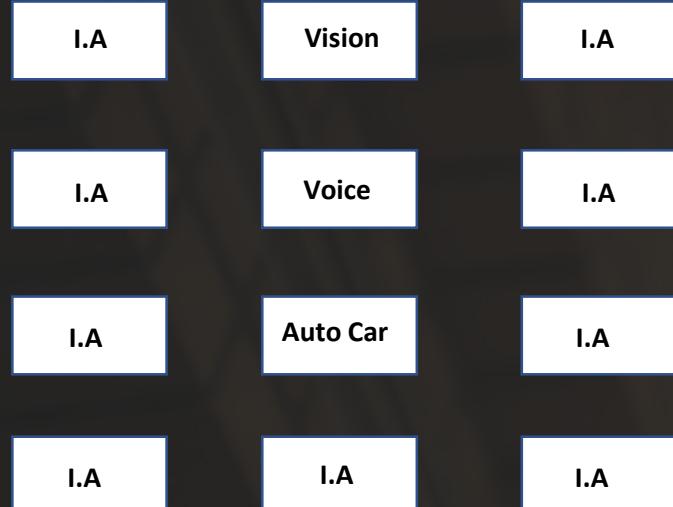




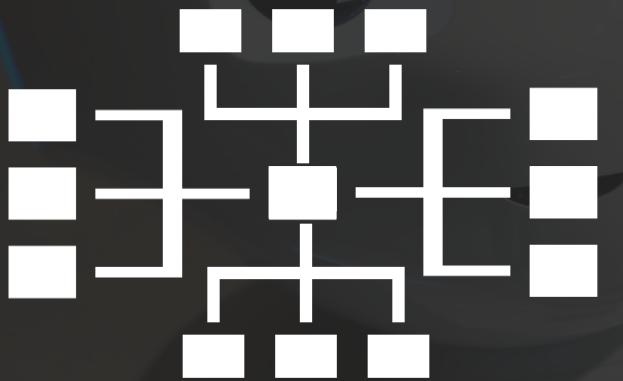


# Tipos de Inteligência Artificial

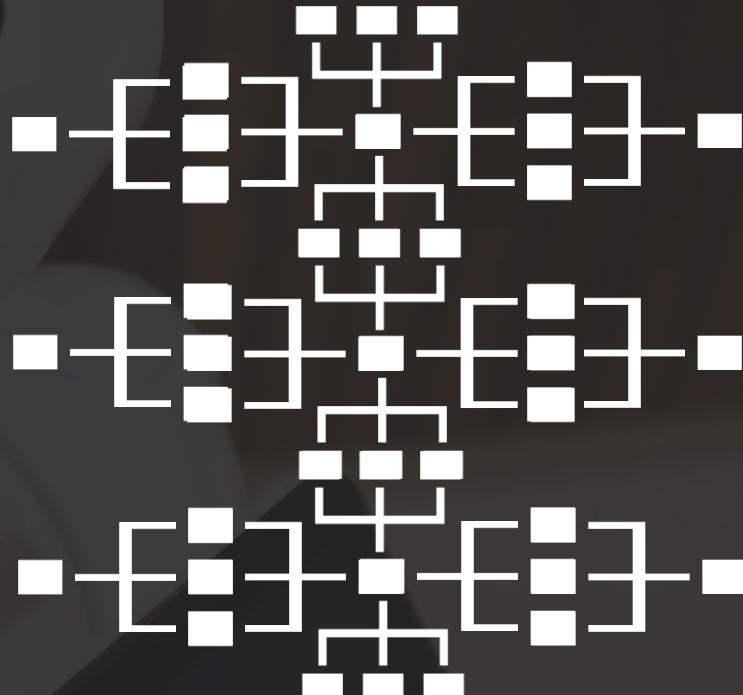
Artificial Narrow Intelligence (**ANI**)

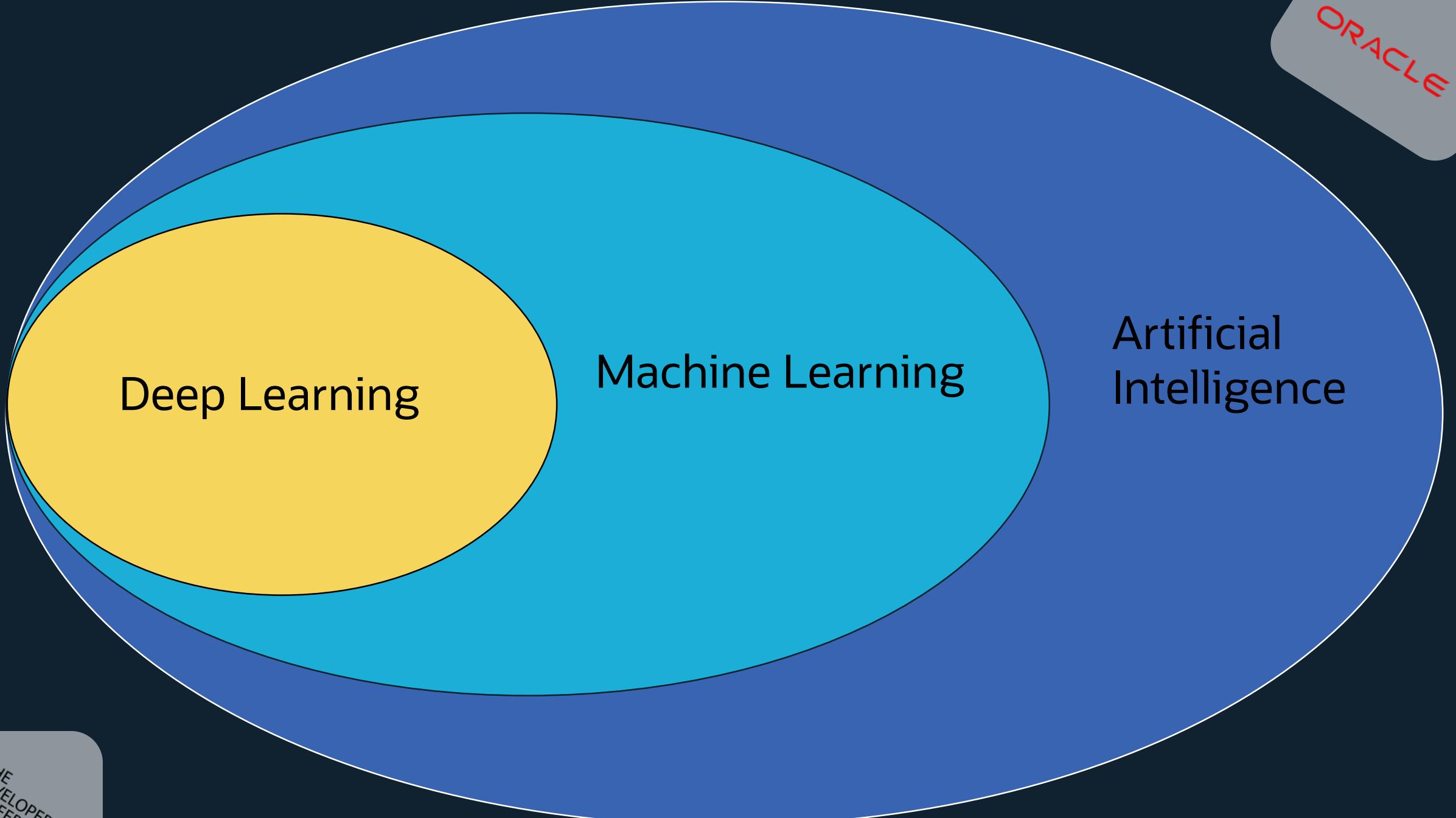


Artificial General Intelligence (**AGI**)



Artificial Superintelligence (**ASI**)



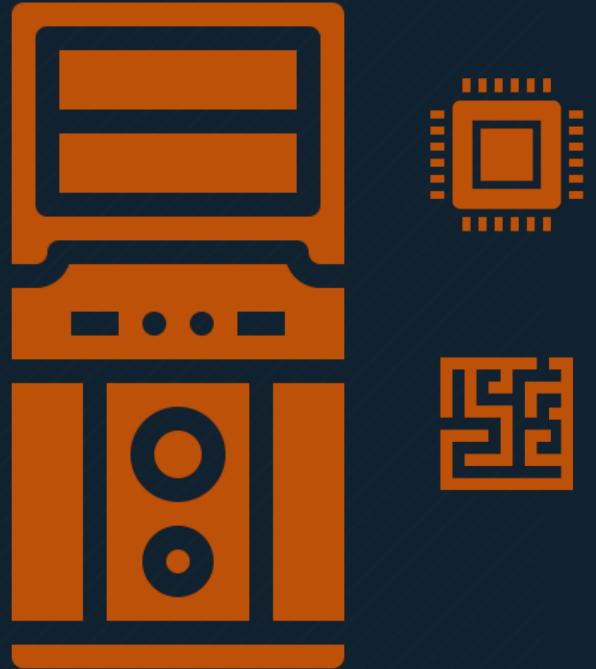


ORACLE



# MACHINE LEARNING

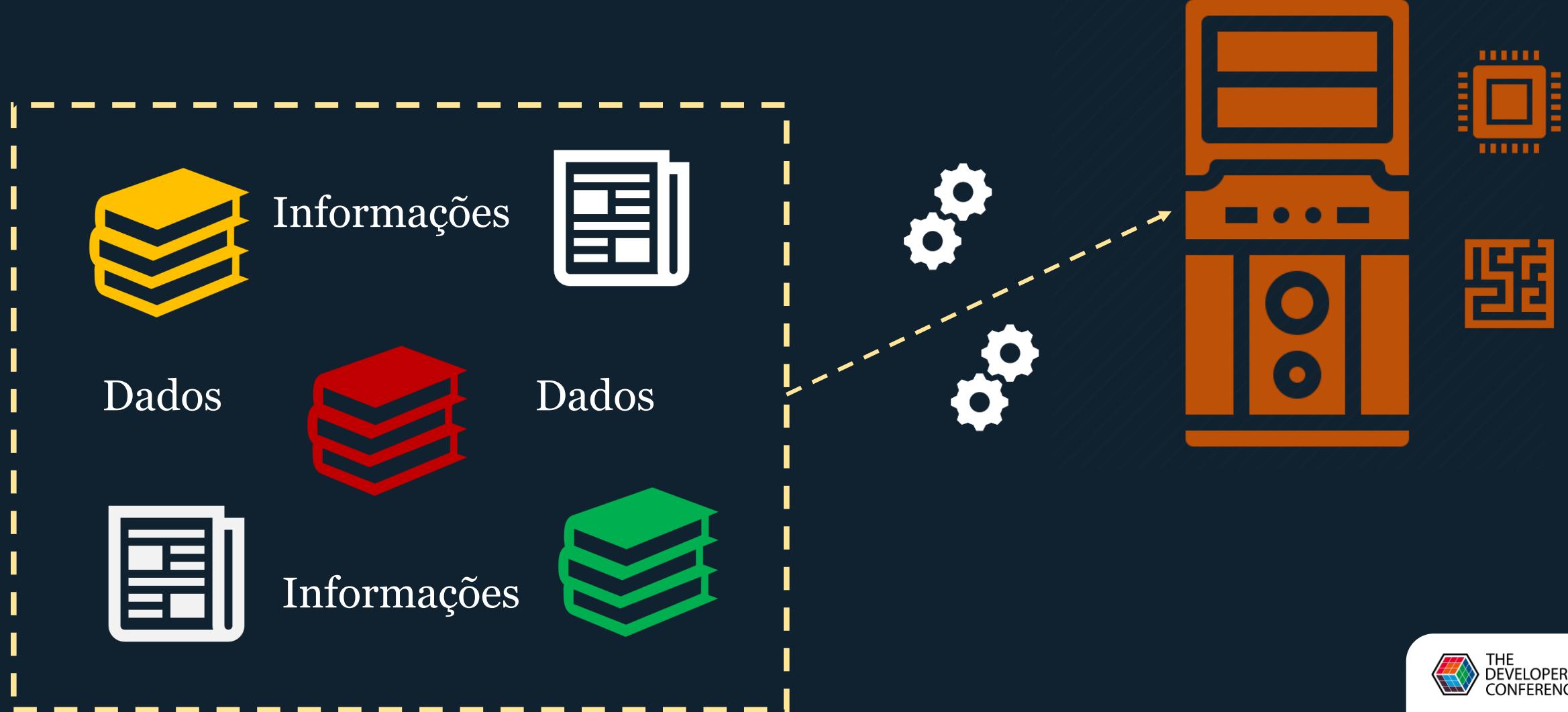
# MACHINE LEARNING - Aprendizado de Máquina



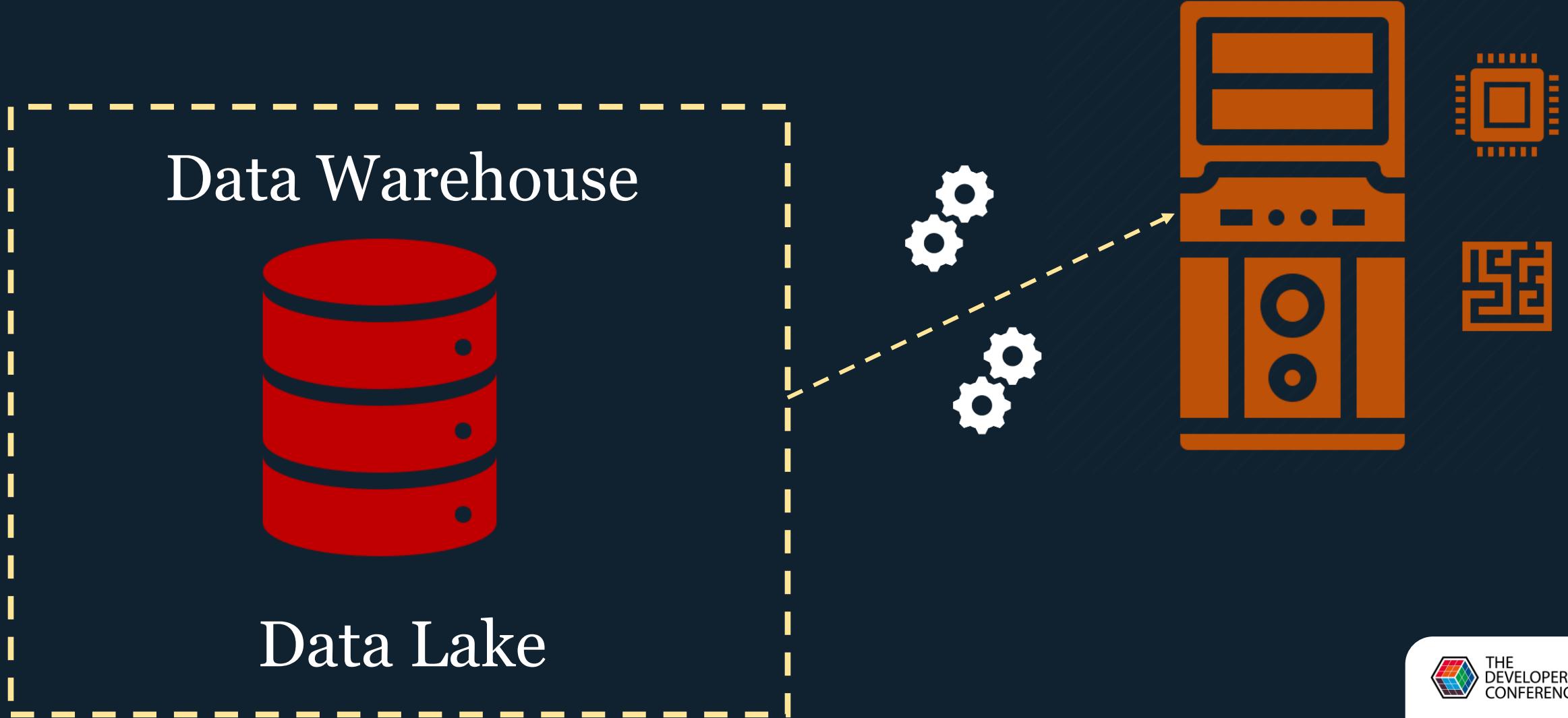
# S T U D Y



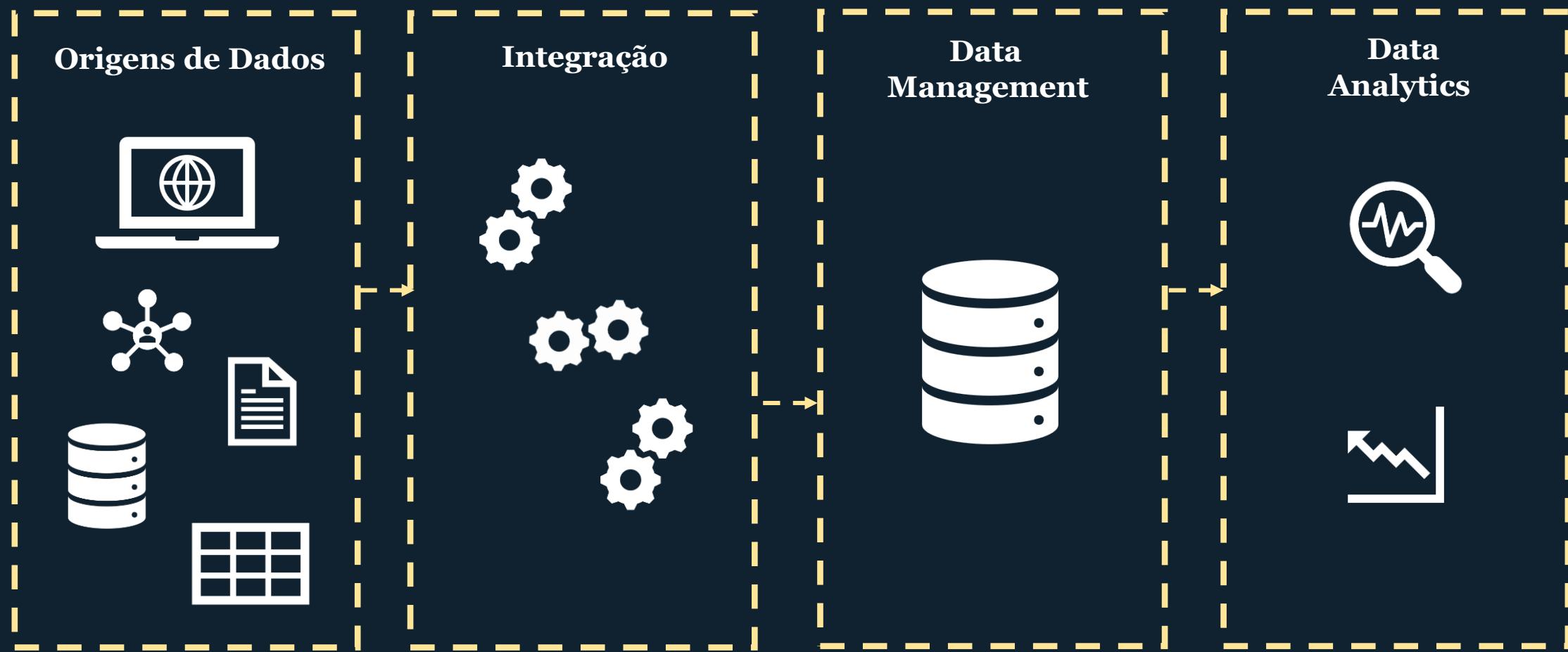
# MACHINE LEARNING - Aprendizado de Máquina



# MACHINE LEARNING - Aprendizado de Máquina



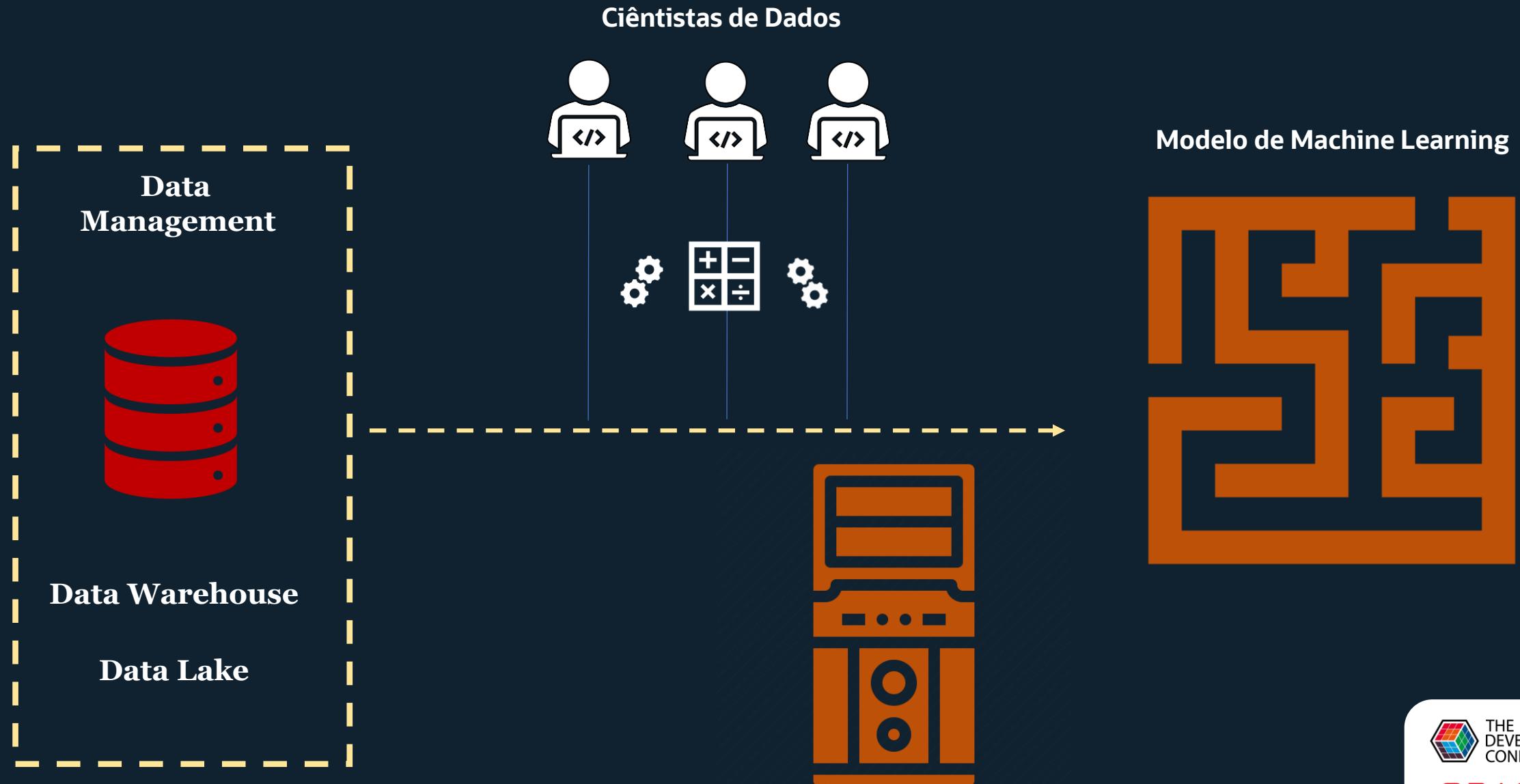
# Plataforma de Dados - Camadas



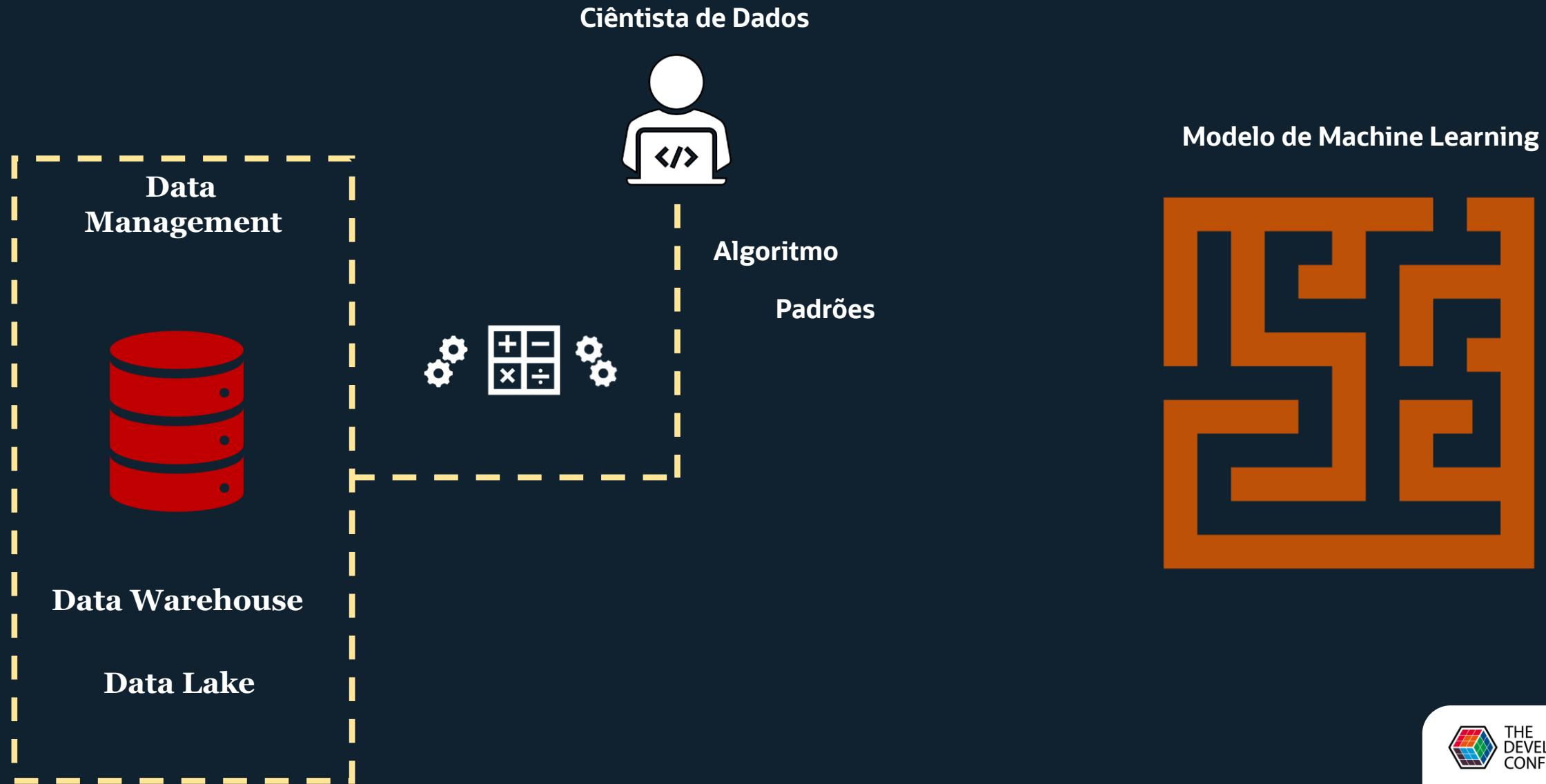
# MACHINE LEARNING - Aprendizado de Máquina



# MACHINE LEARNING - Aprendizado de Máquina



# MACHINE LEARNING - Aprendizado de Máquina



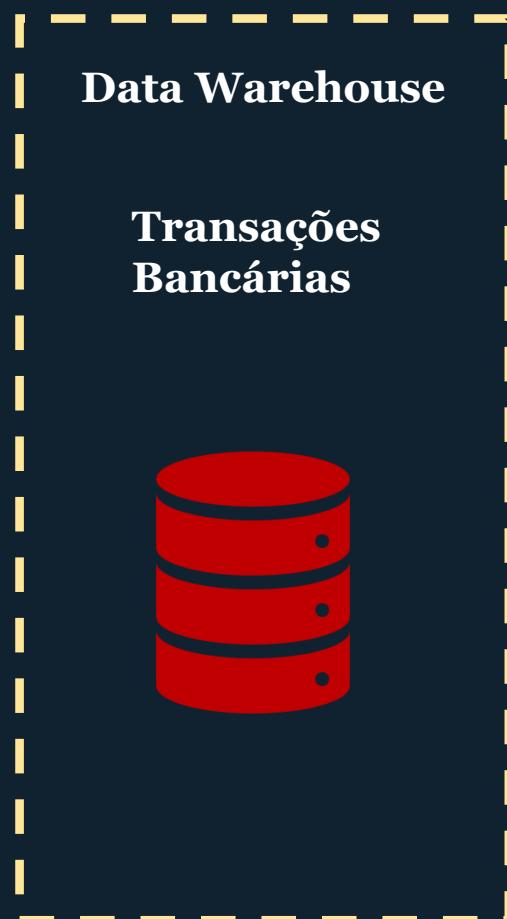
# MACHINE LEARNING - Aprendizado de Máquina

?



# Analisando padrões dos dados

Qual o padrão das fraudulentas?

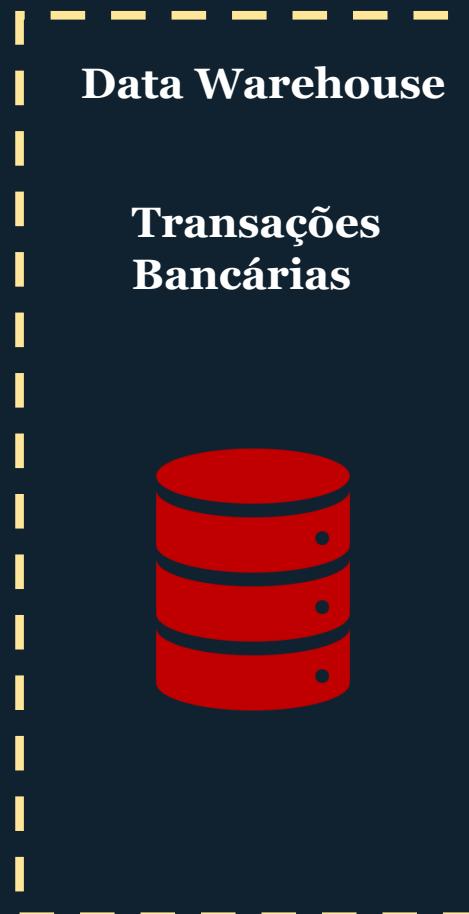


Nome	Valor	Fraudulento
Cleber	R\$ 7.990,00	Sim
Bruna	R\$ 1.291,13	Não
Pedro	R\$ 7.231,84	Sim
Joyce	R\$ 5.110,23	Não

# Analisando padrões dos dados

## MACHINE LEARNING

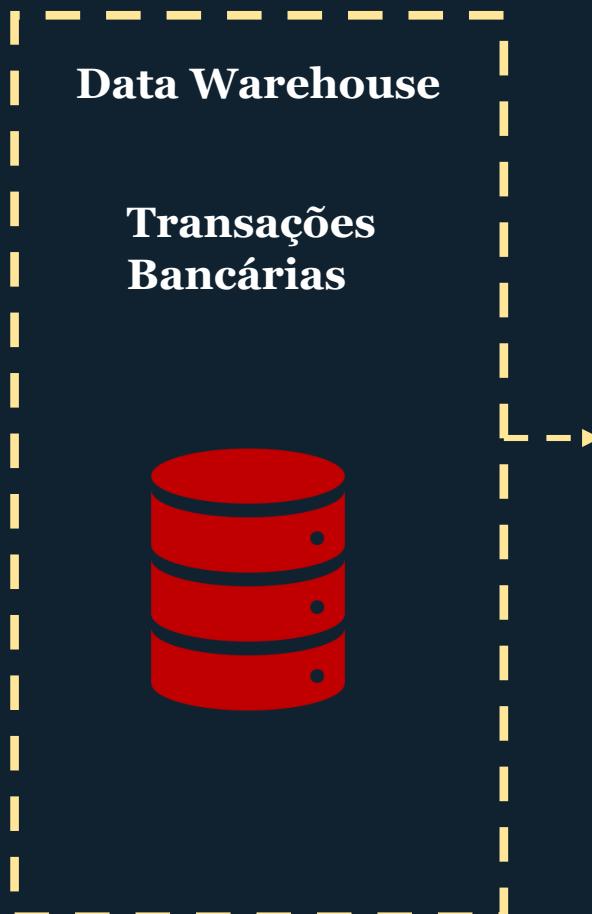
Dados de transações, qual o padrão das fraudulentas?



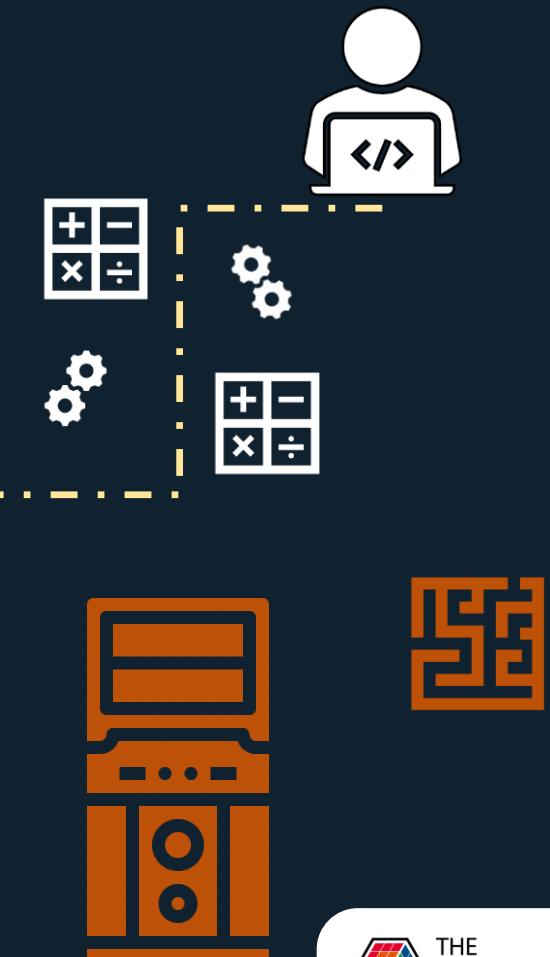
Nome	Valor	Local Emitdo	Local Usado	Idade	Fraudulento
Cleber	R\$ 7.990,00	BR	EUA	30	Sim
Bruna	R\$ 1.291,13	FRA	EUA	37	Não
Pedro	R\$ 7.231,84	BR	EUA	31	Sim
Joyce	R\$ 5.110,23	FRA	EUA	57	Não
Edson	R\$ 728,29	USA	FRA	28	Não
Jack	R\$ 3.123,92	BR	AGR	35	Sim
Felipe	R\$ 2.823,00	BR	EUA	31	Sim
Sabrina	R\$ 7.342,00	EUA	EUA	39	Não
Pedro	R\$ 9.241,00	BR	EUA	36	Não

# Analisando padrões dos dados

Dados de transações, qual o padrão das fraudulentas?



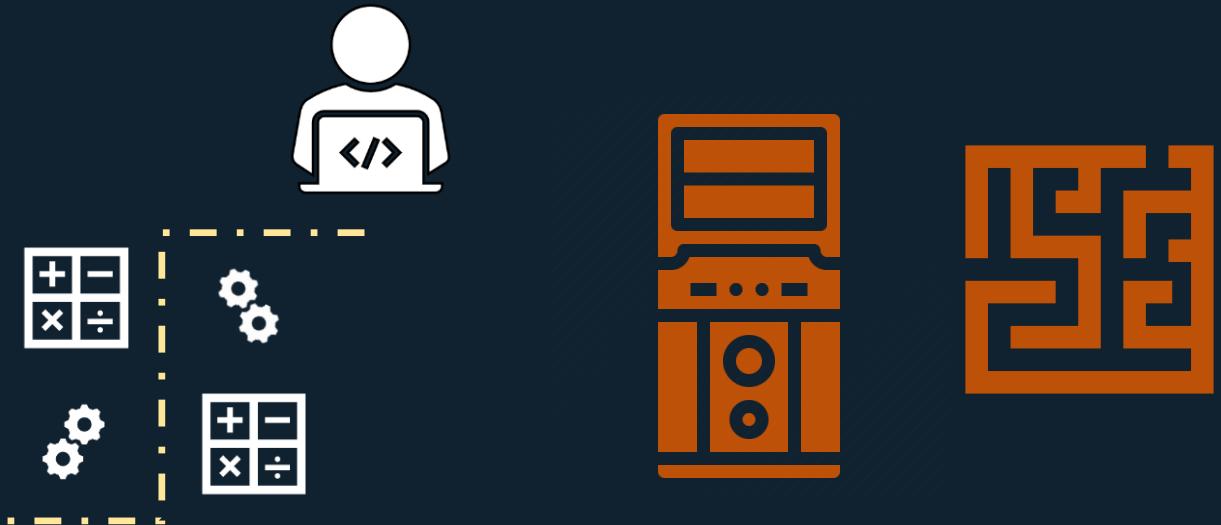
Nome	Valor	Local Emitido	Local Usado	Idade	Fraudulento
Cleber	R\$ 7.990,00	BR	EUA	30	Sim
Bruna	R\$ 1.291,13	FRA	EUA	37	Não
Pedro	R\$ 7.231,84	BR	EUA	31	Sim
Joyce	R\$ 5.110,23	FRA	EUA	57	Não
Edson	R\$ 728,29	USA	FRA	28	Não
Jack	R\$ 3.123,92	BR	AGR	35	Sim
Felipe	R\$ 2.823,00	BR	EUA	31	Sim
Sabrina	R\$ 7.342,00	EUA	EUA	39	Não
Pedro	R\$ 9.241,00	BR	EUA	36	Não



# Analisando padrões dos dados

Dados de transações, qual o padrão das fraudulentas?

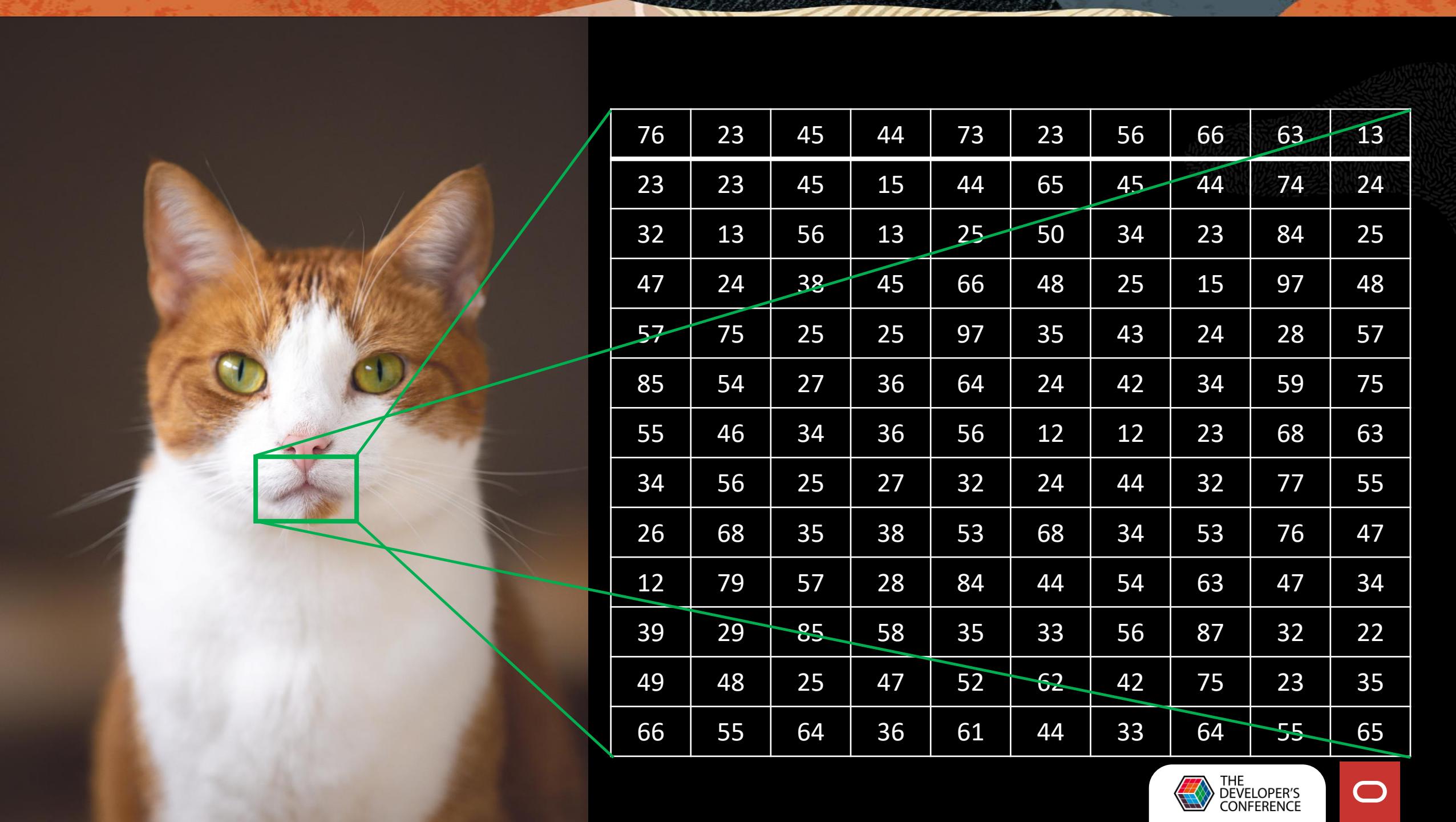
Nome	Valor	Local Emitdo	Local Usado	Idade	Fraudulento
Cleber	R\$ 7.990,00	BR	EUA	30	Sim
Bruna	R\$ 1.291,13	FRA	EUA	37	Não
Pedro	R\$ 7.231,84	BR	EUA	31	Sim
Joyce	R\$ 5.110,23	FRA	EUA	57	Não
Edson	R\$ 728,29	USA	FRA	28	Não
Jack	R\$ 3.123,92	BR	AGR	35	Sim
Felipe	R\$ 2.823,00	BR	EUA	31	Sim
Sabrina	R\$ 7.342,00	EUA	EUA	39	Não
Pedro	R\$ 9.241,00	BR	EUA	36	Não



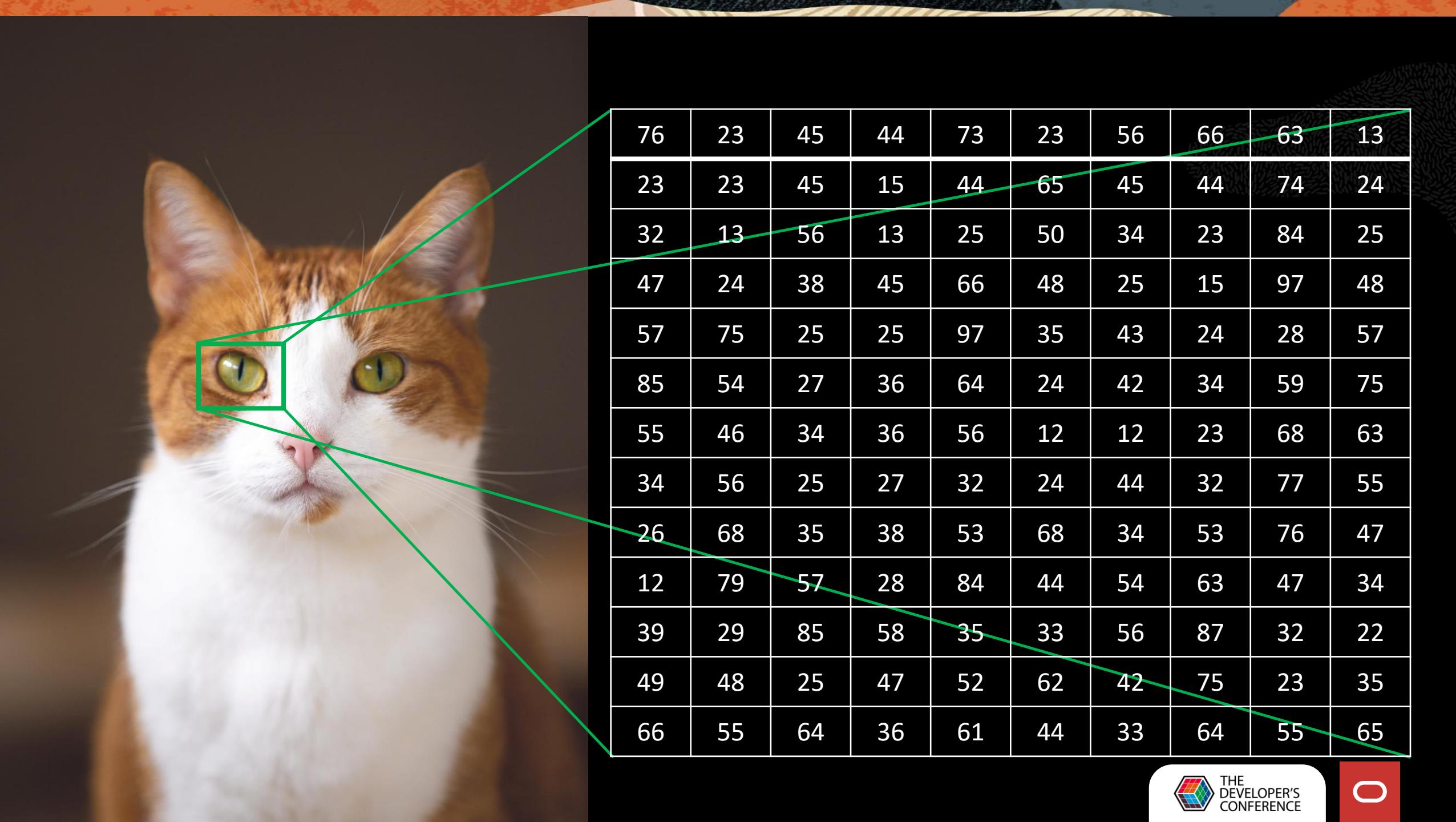
**BR - EUA**

**30 - 40**

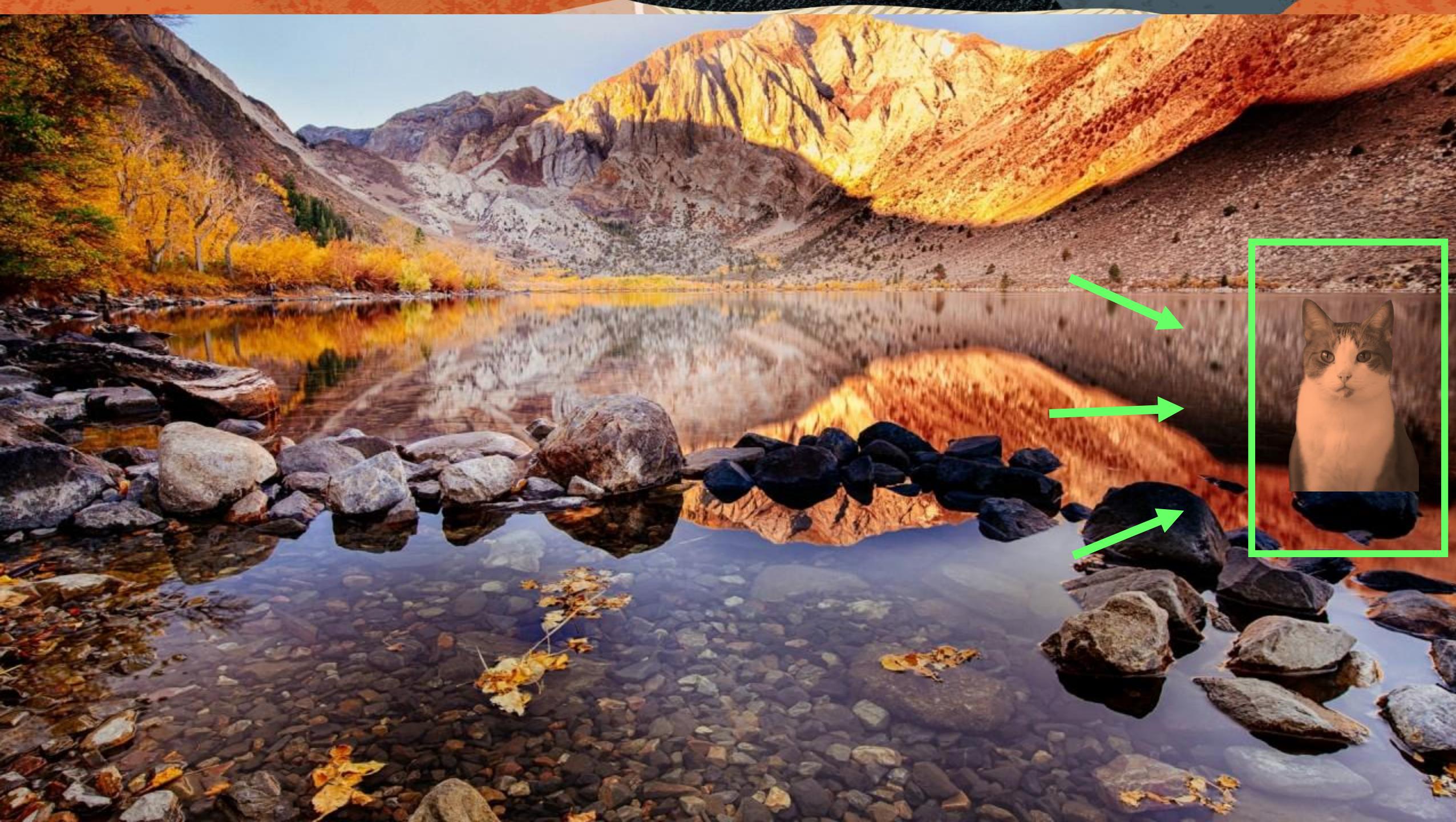
**VALOR > R\$ 1.000,00**



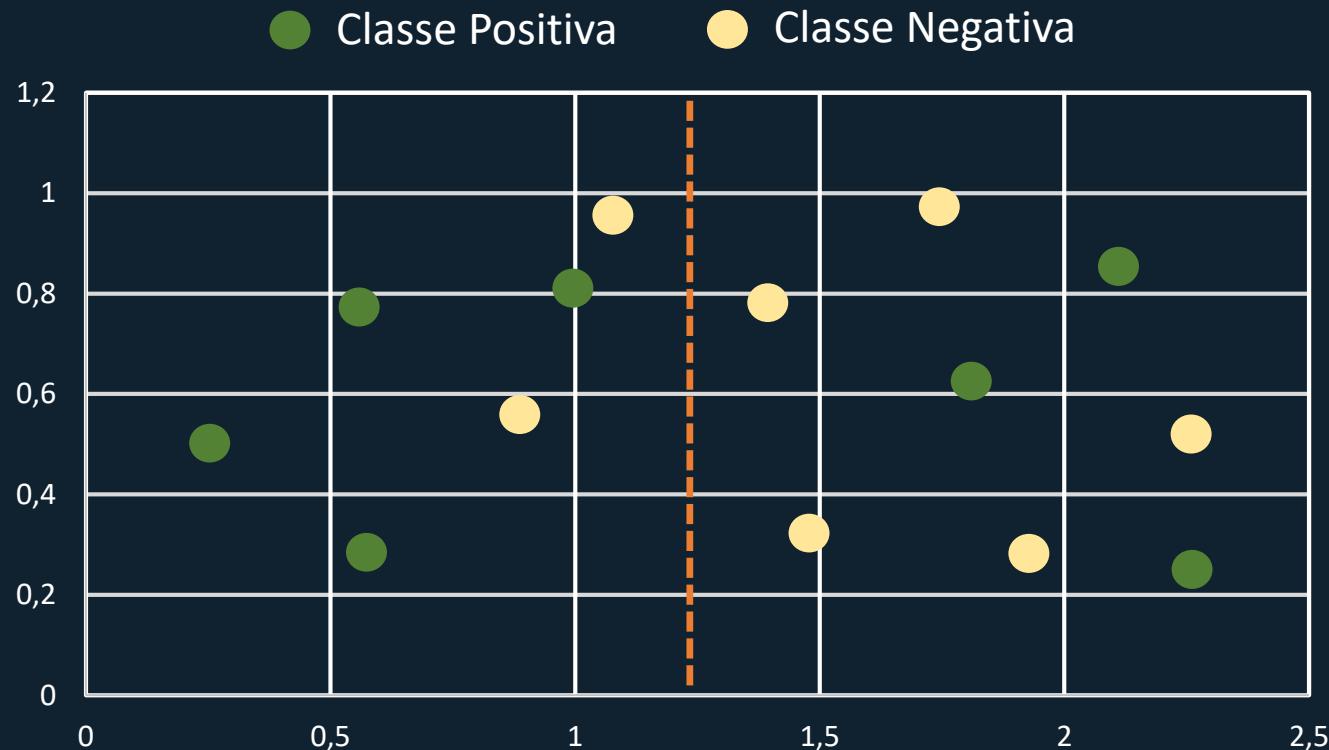
76	23	45	44	73	23	56	66	63	13
23	23	45	15	44	65	45	44	74	24
32	13	56	13	25	50	34	23	84	25
47	24	38	45	66	48	25	15	97	48
57	75	25	25	97	35	43	24	28	57
85	54	27	36	64	24	42	34	59	75
55	46	34	36	56	12	12	23	68	63
34	56	25	27	32	24	44	32	77	55
26	68	35	38	53	68	34	53	76	47
12	79	57	28	84	44	54	63	47	34
39	29	85	58	35	33	56	87	32	22
49	48	25	47	52	62	42	75	23	35
66	55	64	36	61	44	33	64	55	65



76	23	45	44	73	23	56	66	63	13
23	23	45	15	44	65	45	44	74	24
32	13	56	13	25	50	34	23	84	25
47	24	38	45	66	48	25	15	97	48
57	75	25	25	97	35	43	24	28	57
85	54	27	36	64	24	42	34	59	75
55	46	34	36	56	12	12	23	68	63
34	56	25	27	32	24	44	32	77	55
26	68	35	38	53	68	34	53	76	47
12	79	57	28	84	44	54	63	47	34
39	29	85	58	35	33	56	87	32	22
49	48	25	47	52	62	42	75	23	35
66	55	64	36	61	44	33	64	55	65



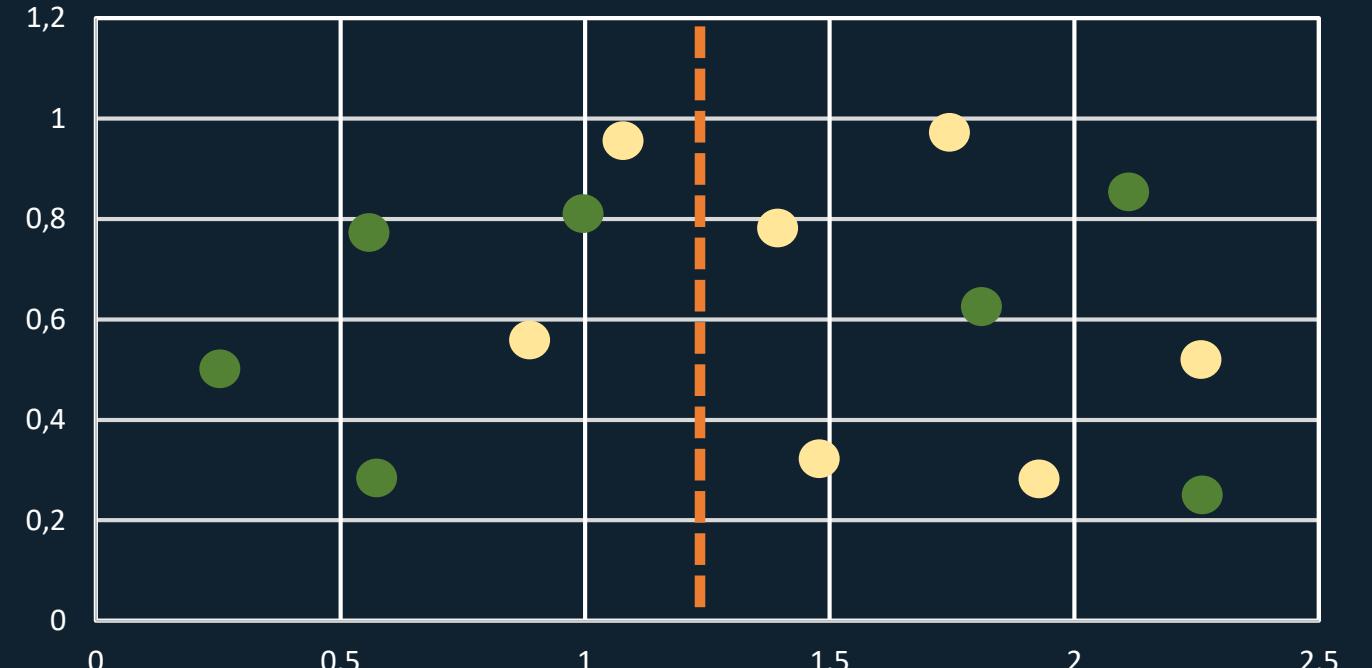
# Matriz de Confusão



**IF < 1.25** ● **Classe Positiva**    **ELSE** ● **Classe Negativa**

IF < 1.25 ● Classe Positiva ELSE ● Classe Negativa

## Matriz de confusão



Valor Previsto		
	Classe Positiva	Classe Negativa
Classe Positiva	4 Verdadeiro Positivo(TP)	3 Falso Negativo(FN)
Classe Negativa	2 Falso Positivo(FP)	5 Verdadeiro Negativo(TN)

**Acurácia:** indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente.

**Precisão:** de todos os dados classificados como positivos, quantos são realmente positivos.

**Recall:** dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas.

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos(TP)} + \text{Verdadeiros Negativos(TN)}}{\text{Total}}$$

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos(TP)}}{\text{Verdadeiros Positivos(TP)} + \text{Falsos Positivos(FP)}}$$

$$\text{Recall} = \frac{\text{Verdadeiros Positivos(TP)}}{\text{Verdadeiros Positivos(TP)} + \text{Falsos Negativos(FN)}}$$

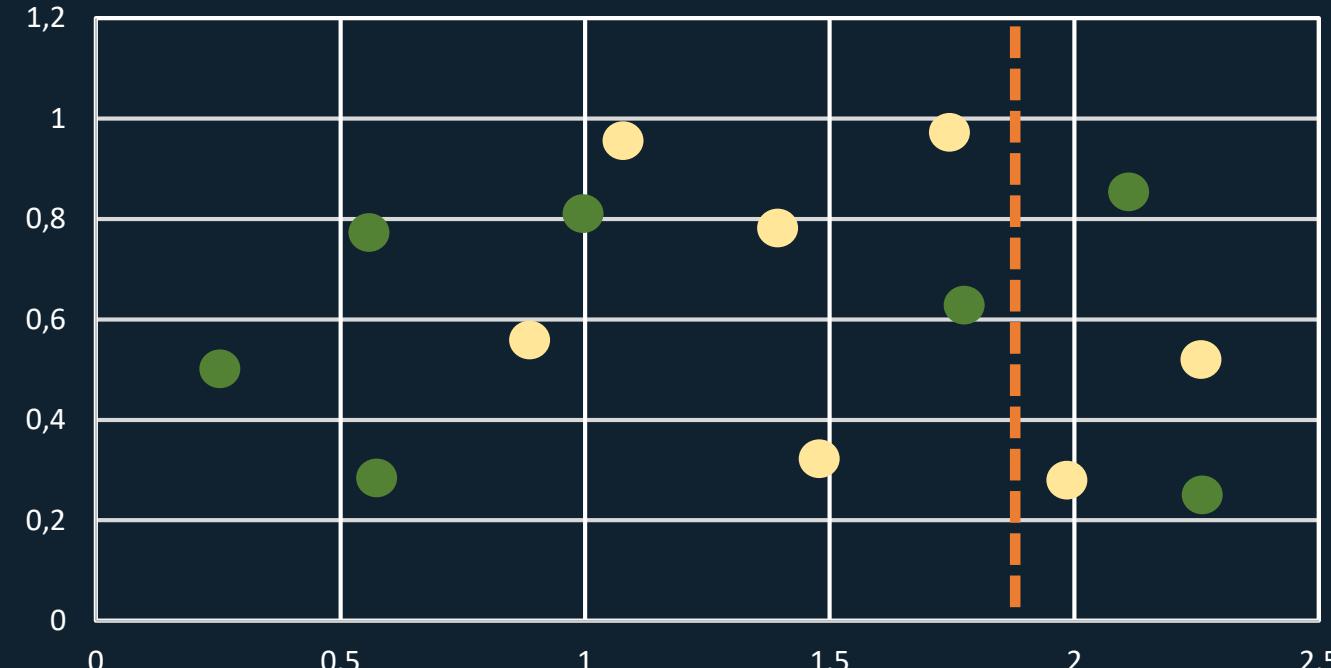
$$\text{Acurácia} = \frac{4 + 5}{14} = 64\%$$

$$\text{Precisão} = \frac{4}{4 + 2} = 66\%$$

$$\text{Recall} = \frac{4}{4 + 3} = 57\%$$

IF < 1.8 ● Classe Positiva ELSE ● Classe Negativa

## Matriz de confusão



		Valor Previsto	
		Classe Positiva	Classe Negativa
Classe Positiva	Verdadeiro Positivo(TP)	5	2
	Falso Positivo(FP)	5	2
Classe Negativa	Falso Negativo(FN)	2	Verdadeiro Negativo(TN)
	Verdadeiro Negativo(TN)	2	Verdadeiro Negativo(TN)

**Acurácia:** indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente.

**Precisão:** de todos os dados classificados como positivos, quantos são realmente positivos.

**Recall:** dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas.

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos}(TP) + \text{Verdadeiros Negativos}(TN)}{\text{Total}}$$

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}(TP)}{\text{Verdadeiros Positivos}(TP) + \text{Falsos Positivos}(FP)}.$$

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}(TP)}{\text{Verdadeiros Positivos}(TP) + \text{Falsos Negativos}(FN)}.$$

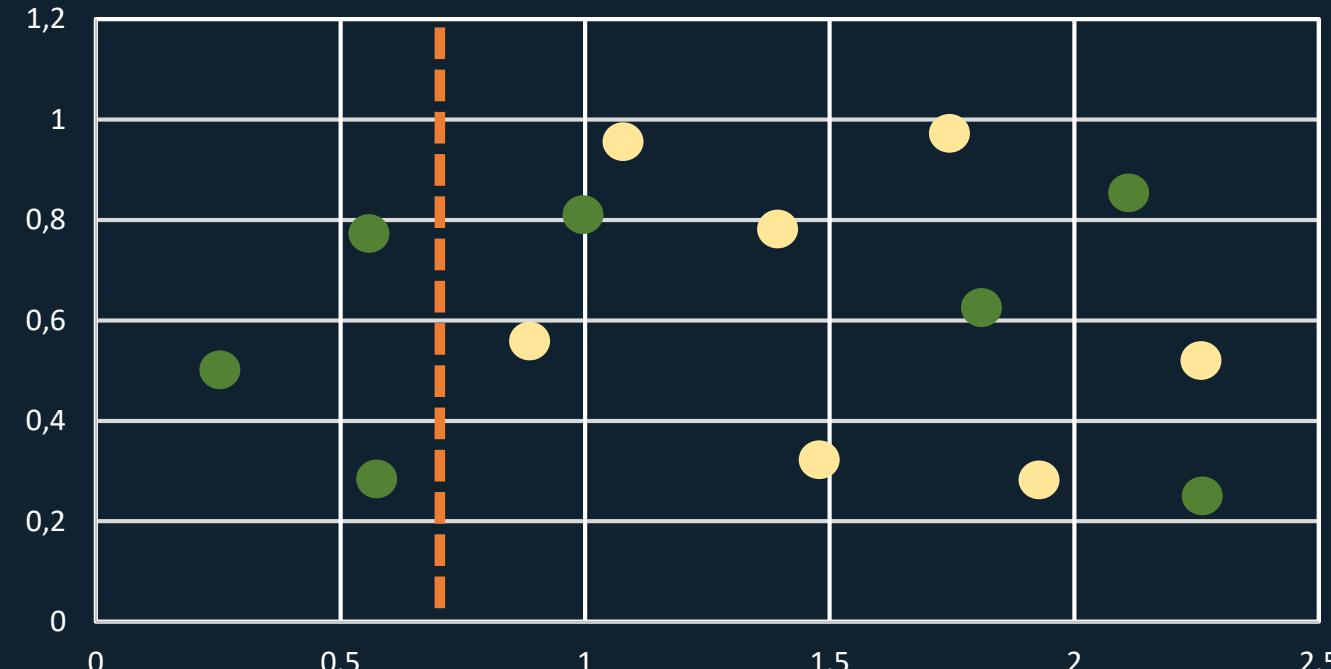
$$\text{Acurácia} = \frac{5 + 2}{14} = 50\%$$

$$\text{Precisão} = \frac{5}{10} = 50\%$$

$$\text{Recall} = \frac{5}{5 + 2} = 71\%$$

IF < 0.6 ● Classe Positiva ELSE ● Classe Negativa

## Matriz de confusão



		Valor Previsto	
		Classe Positiva	Classe Negativa
Classe Positiva	Verdadeiro Positivo(TP)	3	4 Falso Negativo(FN)
	Falso Positivo(FP)	0	7 Verdadeiro Negativo(TN)
Classe Negativa	Verdadeiro Negativo(TN)	7	
	Falso Negativo(FN)	4	

**Acurácia:** indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente.

**Precisão:** de todos os dados classificados como positivos, quantos são realmente positivos.

**Recall:** dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas.

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos}(TP) + \text{Verdadeiros Negativos}(TN)}{\text{Total}}$$

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}(TP)}{\text{Verdadeiros Positivos}(TP) + \text{Falsos Positivos}(FP)}$$

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}(TP)}{\text{Verdadeiros Positivos}(TP) + \text{Falsos Negativos}(FN)}$$

$$\text{Acurácia} = \frac{3 + 7}{14} = 71\%$$

$$\text{Precisão} = \frac{3}{3 + 0} = 100\%$$

$$\text{Recall} = \frac{3}{3 + 7} = 30\%$$

# Hands-on!

# Oracle Data Science

---

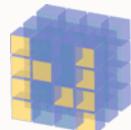
# Oracle Data Science

---



PyTorch

K Keras



NumPy



ANACONDA®

pandas

ATOM

scikit  
learn

SPYDER  
The Scientific Python Development Environment

seaborn



# Oracle Cloud Infrastructure VMs for Data Science

## Pre-configured

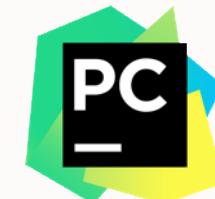
Automated install, update of Linux, supplied packages and Nvidia drivers.

## Any open source library, tool or framework

Comes with over 80 tools, libraries or frameworks. Users can update supplied project, bring their own projects or ISV licenses

## Massive scale

Automated scale out provisioning of multiple VMs to enable training of models on massive data sets that require multiple VMs and GPUs



# O que vamos fazer?

---

Treinar

*Modelo*

Deploy

*Modelo*



# Inovação com dados em nuvem

CIÊNCIA DE DADOS NA PRÁTICA  
COM ORACLE DATA  
SCIENCE

