

# Introdução à IA e watsonX

watsonx.ai

FDC CRI  
Comitê de IA  
27/2/2024



Alan Braz  
Sr Research Developer  
IBM Research



IBM

# IA Generativa já é realidade?!



Write a poem about Xbox



Xbox, oh Xbox, you're my favorite toy  
A source of endless fun, a faithful joy  
From Halo to Gears, you have it all  
A gaming paradise, stand

With your sleek design, a  
controller  
You're the envy of all, a true  
soldier  
Your online multiplayer, a  
Xbox, you're the king, ne



GITHUB COPILOT: CHAT

monalisa

Write unit tests for this function

GitHub Copilot

```
import unittest
import datetime

class TestParseExpenses(unittest.TestCase):
    def test_parse_expenses_with_valid_input(self):
        expenses_string = '2023-01-02 -34.01 USD\n2023-01-03 2.59 DKK
        expected_output = [(datetime.datetime(2023, 1, 2),
                            -34.01, "USD"),
                            (datetime.datetime(2023, 1, 3),
                            2.59, "DKK")]
        self.assertEqual(parse_expenses(expenses_string),
                        expected_output)

    def test_parse_expenses_with_empty_input(self):
        expenses_string = ''
        expected_output = []
        self.assertEqual(parse_expenses(expenses_string),
                        expected_output)

if __name__ == '__main__':
    unittest.main()
```

parse\_expenses.py X addresses.rb X sentiments.ts X

```
1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date, amount, currency)
5     Ignore lines starting with #.
6     Parse the date using datetime.
7     Example expenses_string:
8         2023-01-02 -34.01 USD
9         2023-01-03 2.59 DKK
10        2023-01-03 -2.72 EUR
11
12        expenses = []
13
14        for line in expenses_string.splitlines():
15            if line.startswith("#"):
16                continue
17            date, value, currency = line.split(" ")
18            expenses.append((datetime.datetime.strptime(date, "%Y-%m-%d"),
19                             float(value),
20                             currency))
21
22        return expenses
23
24 expenses_data = ["2023-01-02 -34.01 USD",
25                  "2023-01-03 2.59 DKK",
26                  "2023-01-03 -2.72 EUR"]
```

code assumes that the datetime module is imported. The test cases cover different scenarios such as valid input, empty input, input comments, invalid date format, and invalid value format.

Ask a question or type '/' for commands ➤



[pullrecast.dev](https://pullrecast.dev)

@pullrecast

Inda existe carreira relevante na era da IA generativa?

PULL RECAST

# Mas afinal, o que é IA generativa?

O boom de popularidade começou com  
OpenAI e ChatGPT!

NOVOS EPISÓDIOS SEGUNDAS ÀS 11:33 INSCREVA-SE NO CANAL [PULLRECAST.DEV](https://pullrecast.dev)

12:48 / 1:13:46

Ainda existe carreira relevante na era da IA generativa? – Pull reCast

Pull reCast - Desenvolvendo sua Eminência Técnica  
1.07K subscribers

Subscribed

19 Share Download Clip Save

Impactos do ChatGPT e IA generativa nos negócios e na vida

Impactos do ChatGPT e IA generativa no recrutamento e seleção

Impactos do ChatGPT e IA generativa na arquitetura de soluções

Impactos do ChatGPT e IA generativa em projetos open-source

Impactos do ChatGPT e IA generativa no design de experiências

Impactos do ChatGPT e IA generativa na documentação técnica

Impactos do ChatGPT e IA generativa na segurança da informação

Ainda existe carreira relevante na era da IA generativa?

IMPACTOS DO CHATGPT E IA GENERATIVA NOS NEGÓCIOS E NA VIDA

Impactos do ChatGPT nas carreiras de tecnologia e na vida!

Pull reCast - Desenvolvendo sua Eminência

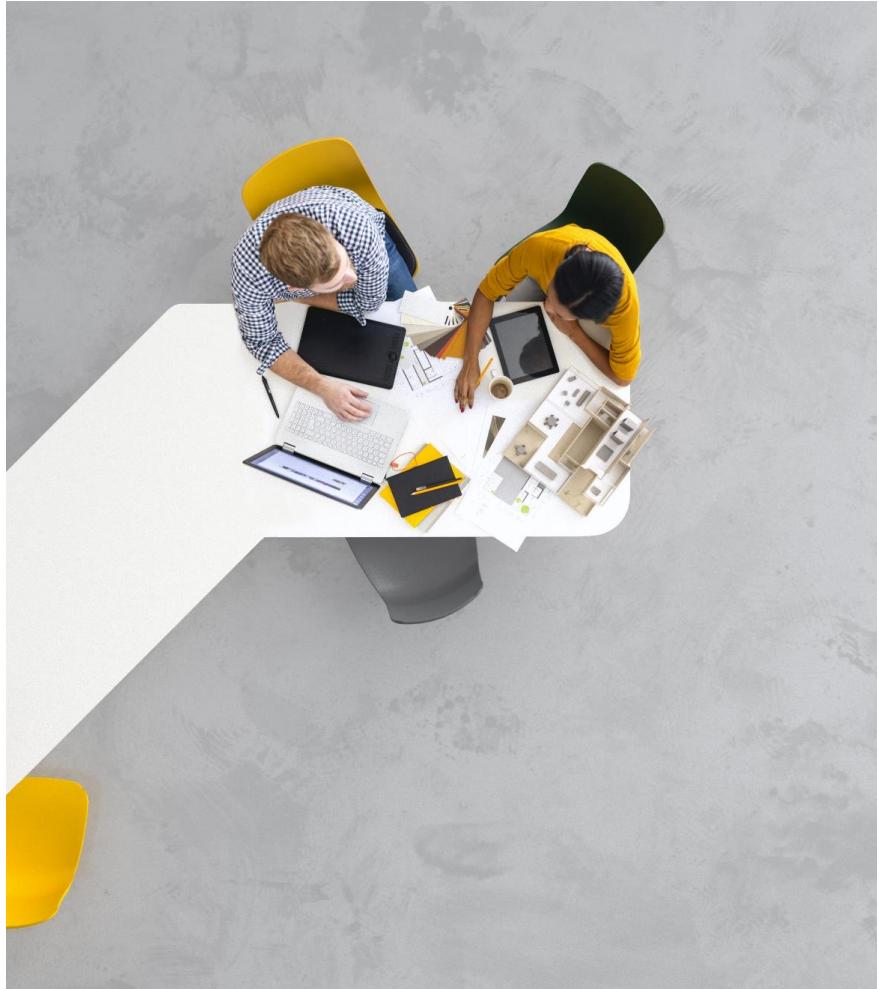
9 videos 35 views Last updated on Aug 1, 2023

Play all Shuffle

Cortes dos episódios discutindo os impactos do ChatGPT e IAs generativas nas carreiras, na produtividade e nas tarefas diárias

- 1 IMPACTOS DO CHATGPT E IA GENERATIVA NOS NEGÓCIOS E NA VIDA 6:50
  - 2 IMPACTOS DO CHATGPT E IA GENERATIVA NO RECRUTAMENTO E SELEÇÃO 4:01
  - 3 IMPACTOS DO CHATGPT E IA GENERATIVA NA ARQUITETURA DE SOLUÇÕES 4:37
  - 4 IMPACTOS DO CHATGPT E IA GENERATIVA EM PROJETOS OPEN-SOURCE 1:42
  - 5 IMPACTOS DO CHATGPT E IA GENERATIVA NO DESIGN DE EXPERIÊNCIAS 3:40
  - 6 IMPACTOS DO CHATGPT E IA GENERATIVA NA DOCUMENTAÇÃO TÉCNICA 20:24
  - 7 IMPACTOS DO CHATGPT E IA GENERATIVA NA DOCUMENTAÇÃO TÉCNICA 8:01
  - 8 IMPACTOS DO CHATGPT E IA GENERATIVA NA SEGURANÇA DA INFORMAÇÃO 3:03
  - 9 Ainda existe carreira relevante na era da IA generativa? 1:13:47
- Ainda existe carreira relevante na era da IA generativa? – Pull reCast

# Introdução à IA



# What is AI?

*AI refers to the ability of computer systems to attempt to mimic the problem-solving and decision-making capabilities of the human mind.*



Computer vision



Data science

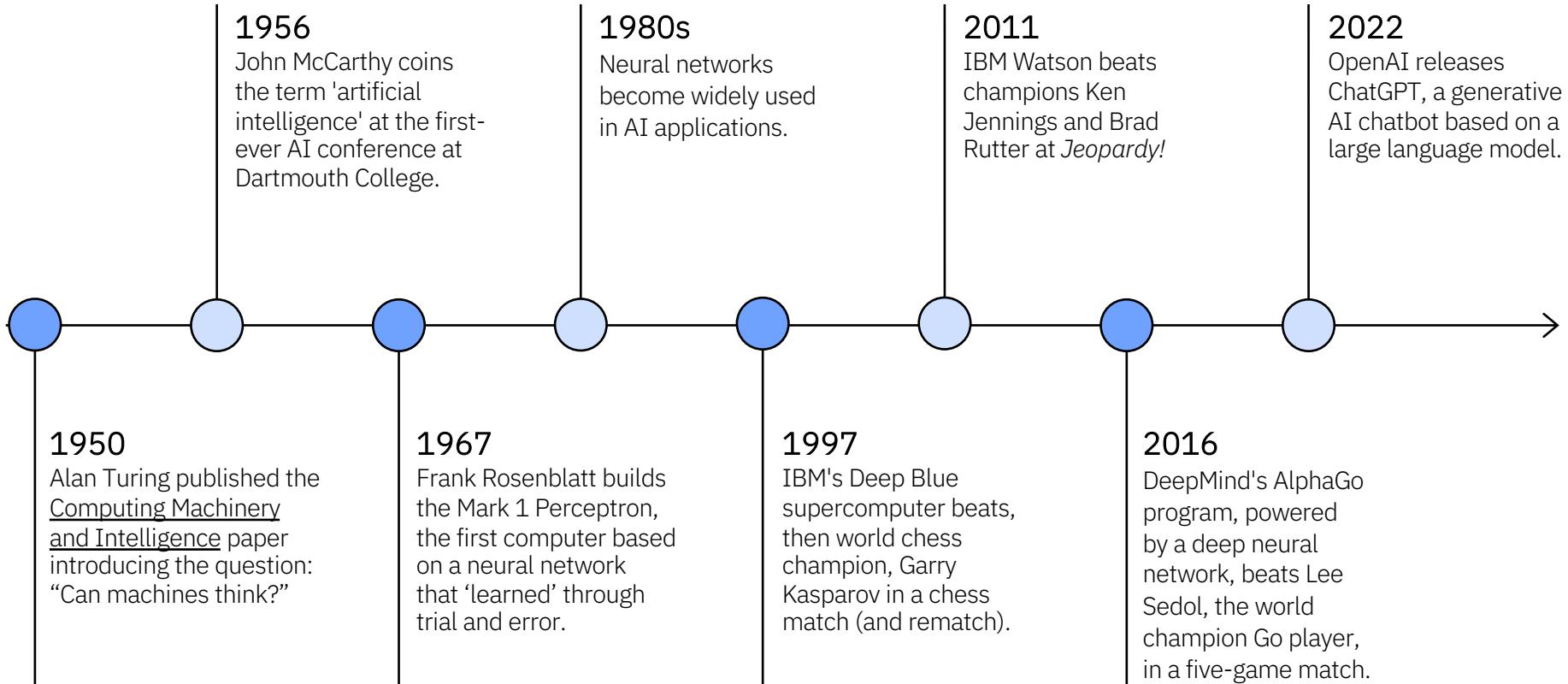


Natural  
language  
processing  
(NLP)



Robotics

# AI milestones



# Deep Blue 1997



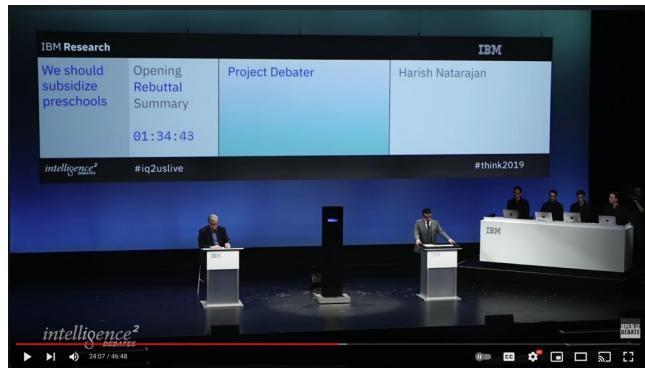
# Ei! 2013/2014



# Watson 2011



# Project Debater 2019



<https://research.ibm.com/interactive/project-debater/film/>  
[https://youtube.com/watch?v=3\\_yy0dnIc58](https://youtube.com/watch?v=3_yy0dnIc58)

Análise de sentimento durante a Copa usando Big Data

Apresentação gravada no TDC 2014

IBM lança o primeiro projeto de análise de sentimentos em português para grande volume de dados

# Artificial Intelligence (AI)

*Human intelligence exhibited by machines*



AI can be defined as a technique that enables machines to mimic cognitive functions associated with human minds – cognitive functions include all aspects of learning, reasoning, perceiving, and problem solving.

## Machine Learning (ML)

*Systems that learn from historical data*



ML-based systems are trained on historical data to uncover patterns. Users provide inputs to the ML system, which then applies these inputs to the discovered patterns and generates corresponding outputs.

## Deep Learning (DL)

*ML technique that mimics human brain function*



DL is a subset of ML, using multiple layers of neural networks, which are interconnected nodes, which work together to process information. DL is well suited to complex applications, like image and speech recognition.

## Foundation Model

*Generative AI systems*



AI model built using a specific kind of neural network architecture, called a transformer, which is designed to generate sequences of related data elements (for example, like a sentence).

1950's

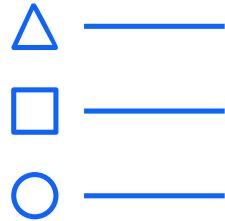
1980's

2010's

2020's



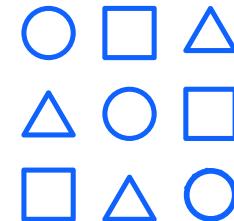
## Rules-based systems



*programmed with a series  
of [instructions](#) & logical rules*

*for example: “a triangle has three sides”*

## Machine learning systems

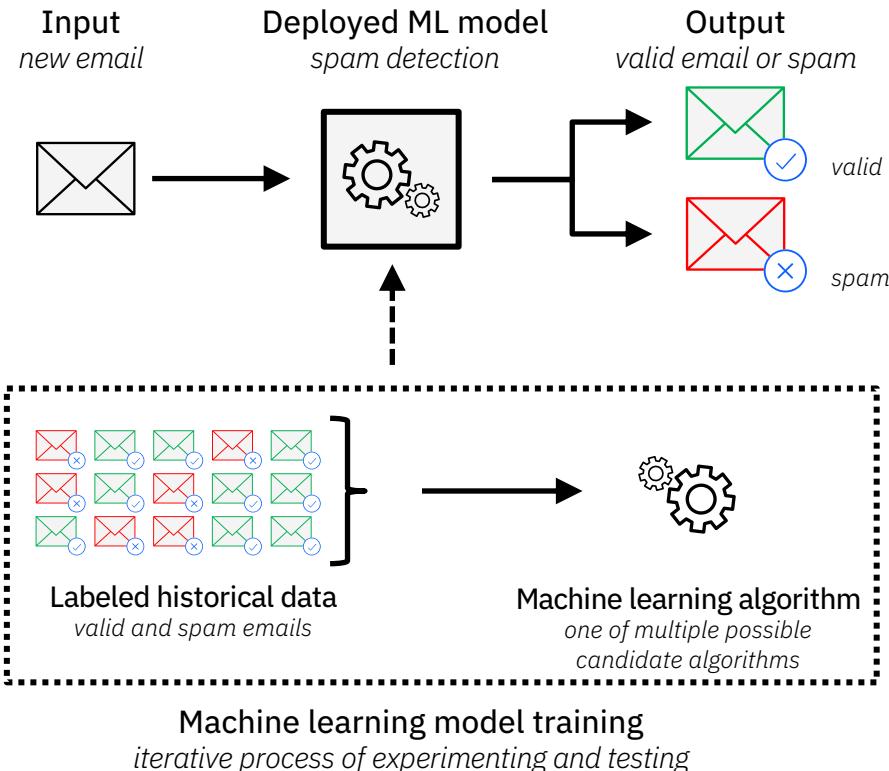


*trained with a  
series of [examples](#)*

*for example: “here are  
pictures of many triangles”*

# What is a machine learning (ML) model?

*Example: Spam detection for email*



# Machine learning (ML) types

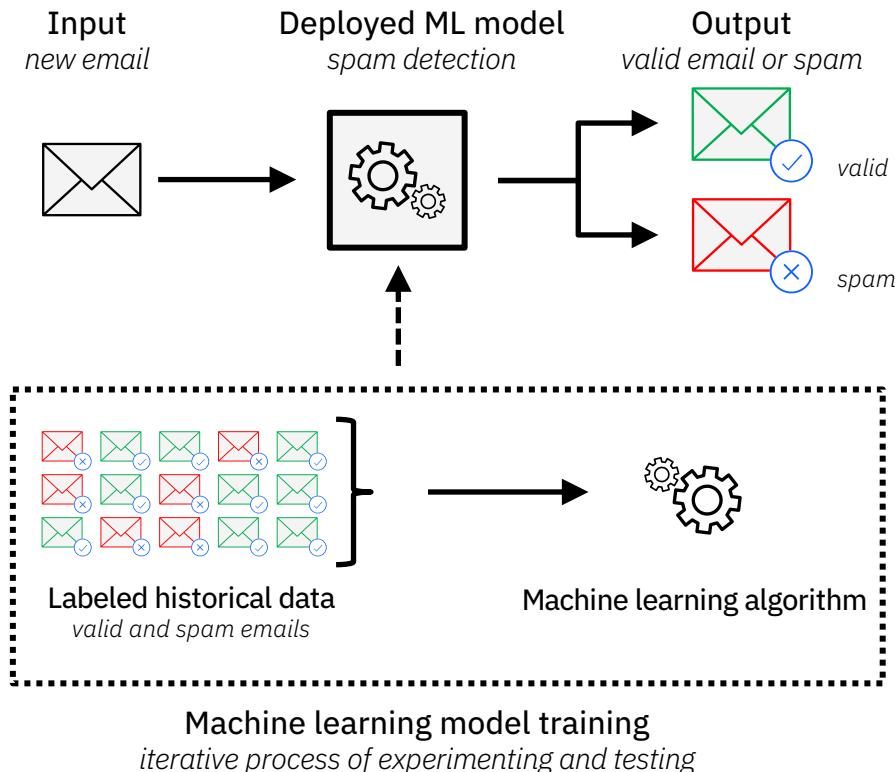
## *Classification models*

Classification models assign labels to model inputs or assign them to specific categories.

Common use cases include:

- Fraud detection: predict whether a transaction is fraudulent based on patterns in the data
- Sentiment analysis: classify text as positive, negative, or neutral
- Medical diagnosis: assign a disease label to a patient's case, based on symptoms and medical history
- Image recognition: recognize objects or identify people based on visible features and characteristics

*Example: Spam detection for email*



# Machine learning (ML) types

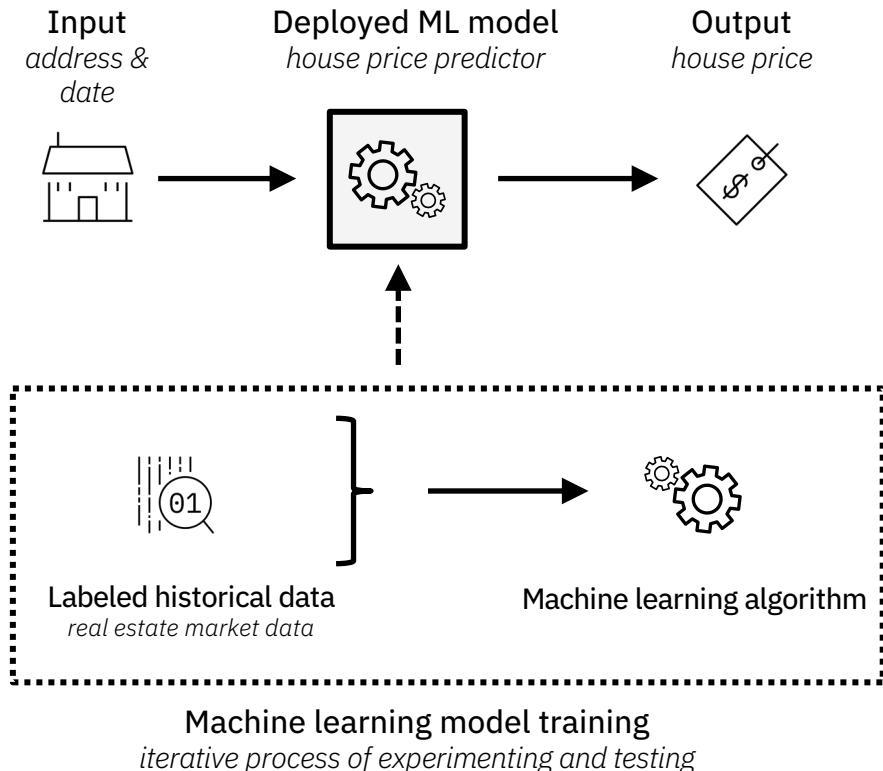
## Regression models

Regression models make predictions based on the model input.

Common use cases include:

- Stock market analysis: securities price prediction based on historical data or news events
- Sales: forecasting based on historical data or market trends
- Healthcare: predict patient outcomes based on factors such as age, gender, medical history, or treatment plans
- Customer behavior analysis: predict future customer purchasing patterns based on demographic data, past purchase history, and advertising campaigns

*Example: House price prediction*



# Machine learning (ML) types

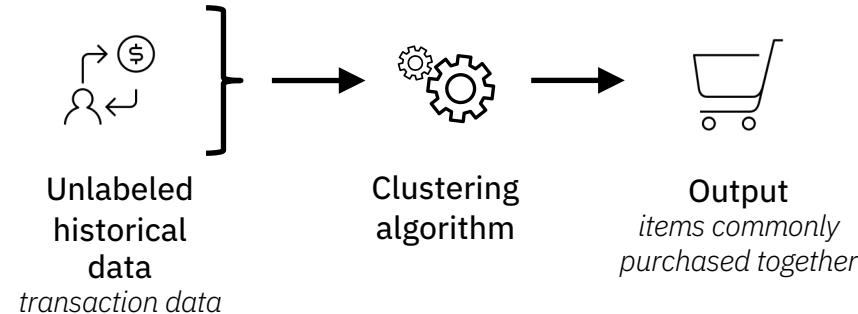
## Clustering models

Clustering models identifies distinct groupings of individual data points that share common characteristics within a larger data set.

Common use cases include:

- Customer segmentation: group customers based on similar preference, behaviors, and demographics
- Genetic analysis: group genes with similar functions or processes
- Social network analysis: identify communities or groups within a social network
- Market basket analysis: identify items that are commonly purchased together

*Example: Market basket analysis*



# Machine learning (ML) types

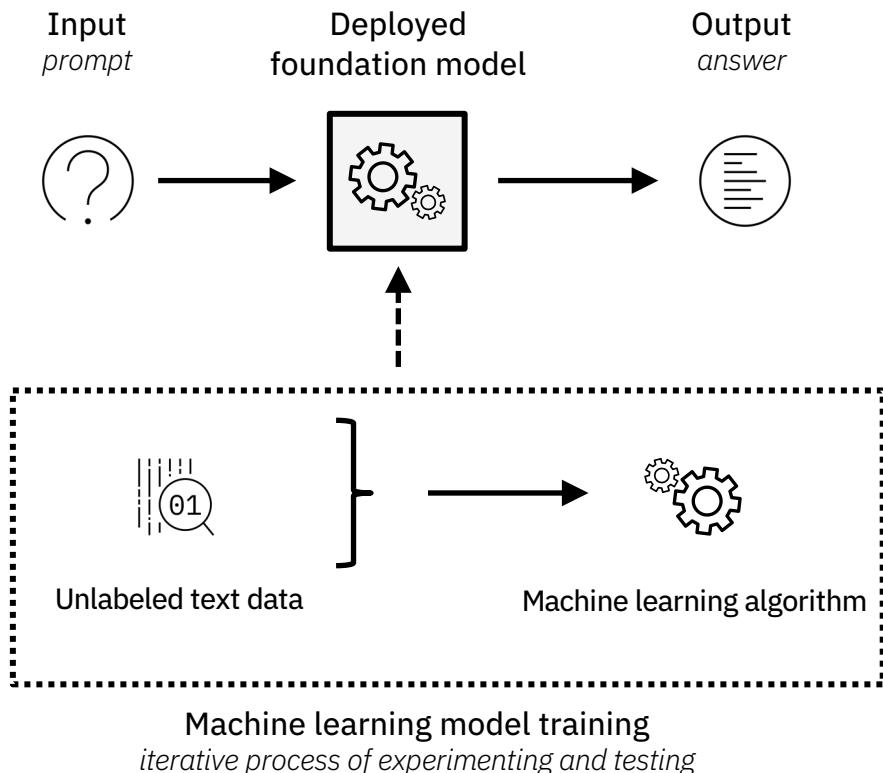
## Generative models

Generative models create new data in response to input requests (prompts).

Common use cases include:

- Text generation: generate new text in response to a prompt, summarizing text, or writing a lengthy essay
- Code generation: generate computer code based on a textual description of the proposed program
- Image generation: create images based on the prompt

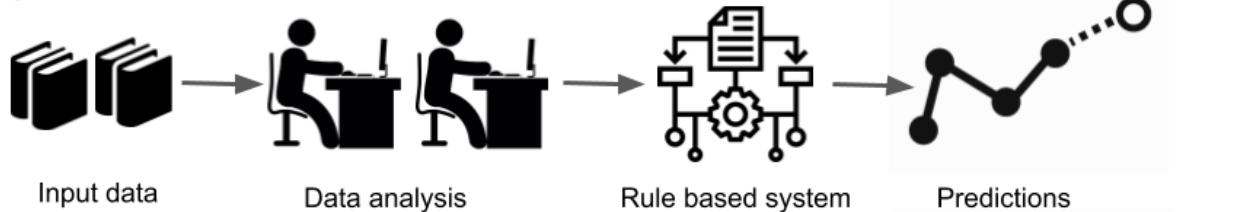
*Example: Text generation*



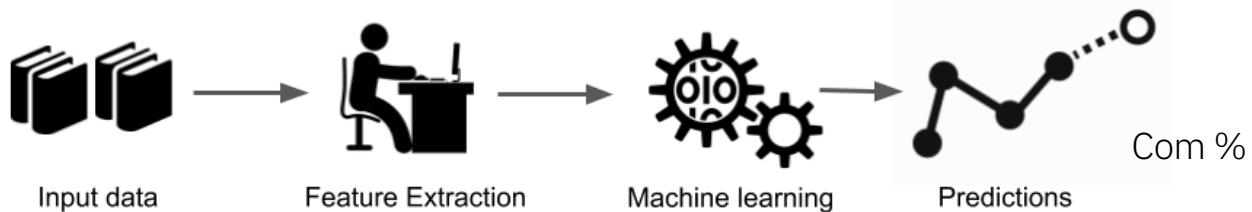


# Paradigmas

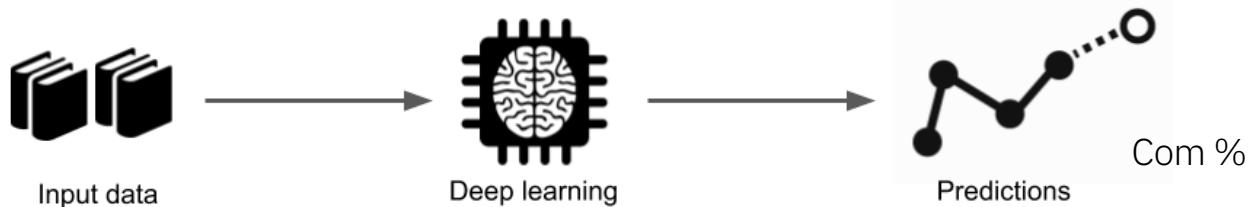
Poucos e  
normalmente estruturados



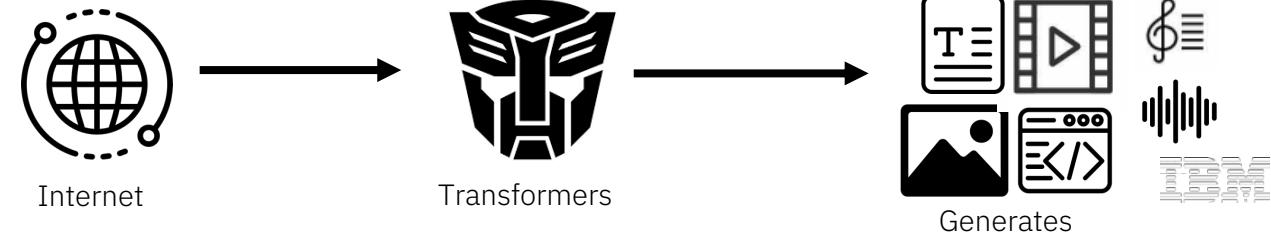
Milhares,  
Estruturados ou não  
Anotados!



Milhões  
Estruturados ou não  
Não anotados!



Bilhões/Trilhões  
Não Estruturados  
Não anotados!



# Large language models (LLMs)

LLMs are machine learning models that have been trained on large quantities of unlabeled text using self-supervised learning and can perform a variety of natural language processing tasks .

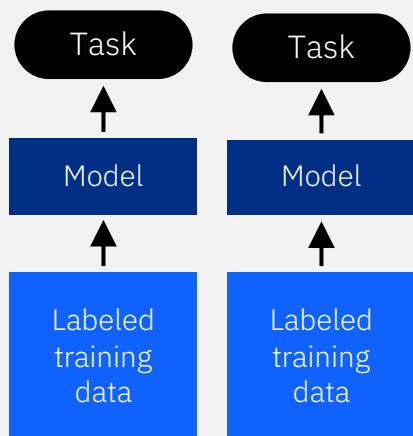
ChatGPT is an example of a generative AI chatbot developed using an LLM.



## What's next with foundation models:

Conventional AI

1000s – 100000s labeled  
data points / task



Zero-shot prompting

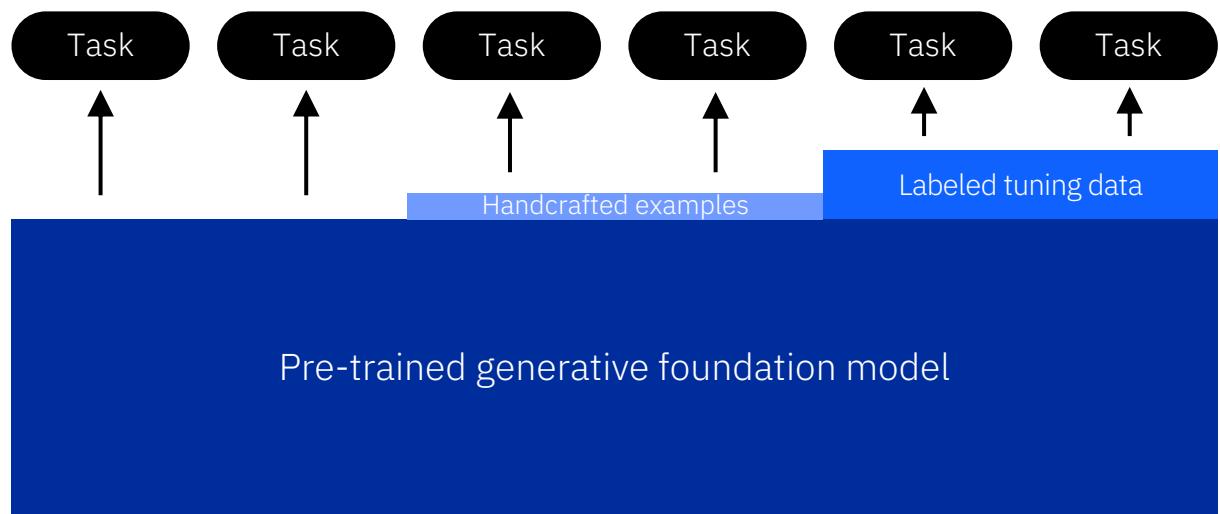
No labeled data needed

Few-shot prompting

1-10 labeled data points /  
task

Data-driven tuning

100s – 1000s labeled  
data points / task



## Inteligência Artificial (IA)

Inteligência humana simulada por máquinas

IA pode ser definida como uma técnica que permite que as máquinas imitem as funções cognitivas associadas às mentes humanas - as funções cognitivas incluem todos os aspectos da aprendizagem, raciocínio, percepção e resolução de problemas.

**1950**

Alan Turing publicou o artigo Computing Machinery and Intelligence introduzindo a pergunta: "Máquinas podem pensar?"

**1956**

John McCarthy cunhou o termo 'inteligência artificial' na primeira conferência de IA no Dartmouth College.

### Machine Learning (ML)

Sistemas que aprendem com dados históricos

Aprendizado supervisionado

1Ks a 1Ms de exemplos de dados rotulados

Cada tarefa um modelo

Classificação, Regressão, Árvores de decisão

"Aprendizado de Máquina": Os sistemas baseados em ML são treinados em dados históricos para descobrir padrões. Os usuários fornecem entradas para o sistema ML, que aplica essas entradas aos padrões descobertos e gera as saídas correspondentes.

### Deep Learning (DL)

Técnica de ML que imita a função do cérebro humano

Aprendizado semi- ou não-supervisionado

1Ms de exemplos de dados rotulados e não rotulados

Imagens e sons

"Aprendizado Profundo": DL é um subconjunto de ML, usando várias camadas de redes neurais, que são nós interconectados, que trabalham juntos para processar informações. DL é adequado para aplicativos complexos, como reconhecimento de imagem e fala.

### Foundation Model

Sistemas de IA generativos

Aprendizado não supervisionado  
Bilhões de dados não rotulados  
Grandes modelos de linguagem LLM  
Modelo "base" pré-treinado de forma genérica  
GPT-3: 1024 GPUs, 34 dias, \$4.6M

Modelo de IA construído usando um tipo específico de arquitetura de rede neural, chamado Transformer, projetado para gerar sequências de elementos de dados relacionados (por exemplo, como uma frase).

Fine-tuning:

Zero-shot prompting: No labeled data

Few-shot prompting: 1-10 labeled data

Data-driven tuning: 100s – 1000s labeled

1950's

1980's

2010's

2020's

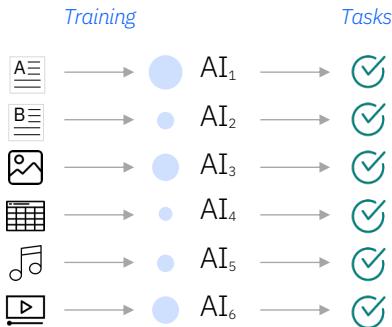
<https://pullrecast.dev/images/resumo-ia.png>

## The impact of generative AI |

The opportunity

# Foundation models establish a new paradigm for AI capabilities

## Traditional AI models



- Individual siloed models
- Require task specific training
- Lots of human supervised training

## Foundation models

### External data



+ Enterprise  
proprietary data

Massive  
foundation  
model

+ Enterprise  
proprietary data

Fine-tuned  
model

- Massive multi-tasking model
- Adaptable with minimized training
- Pre-trained unsupervised learning

Prompting

Q&A

Translation

Sentiment  
analysis

Code  
generation

...

...

## Enhanced capabilities

- Summarization
- Conversational knowledge
- Content creation
- Code co-creation

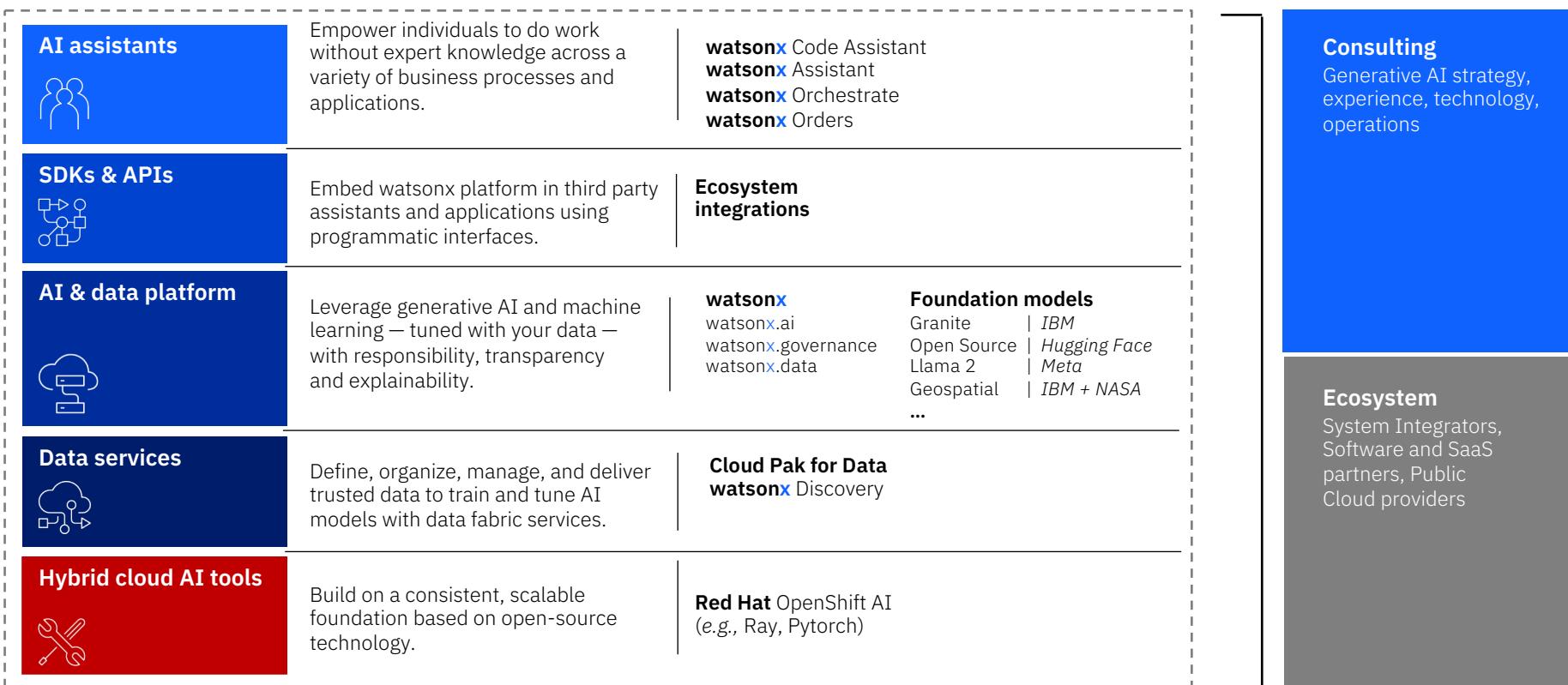
## Key advantages

- Lower upfront costs through less labeling
- Faster deployment through fine tuning
- Equal or better accuracy for multiple use cases
- Incremental revenue through better performance

up to **70% reduction**  
in certain NLP tasks

## What IBM offers

# IBM's generative AI technology and expertise



What IBM offers

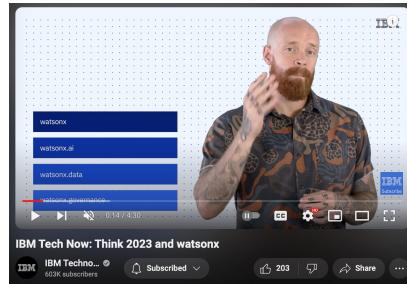
Introducing...

# watsonx

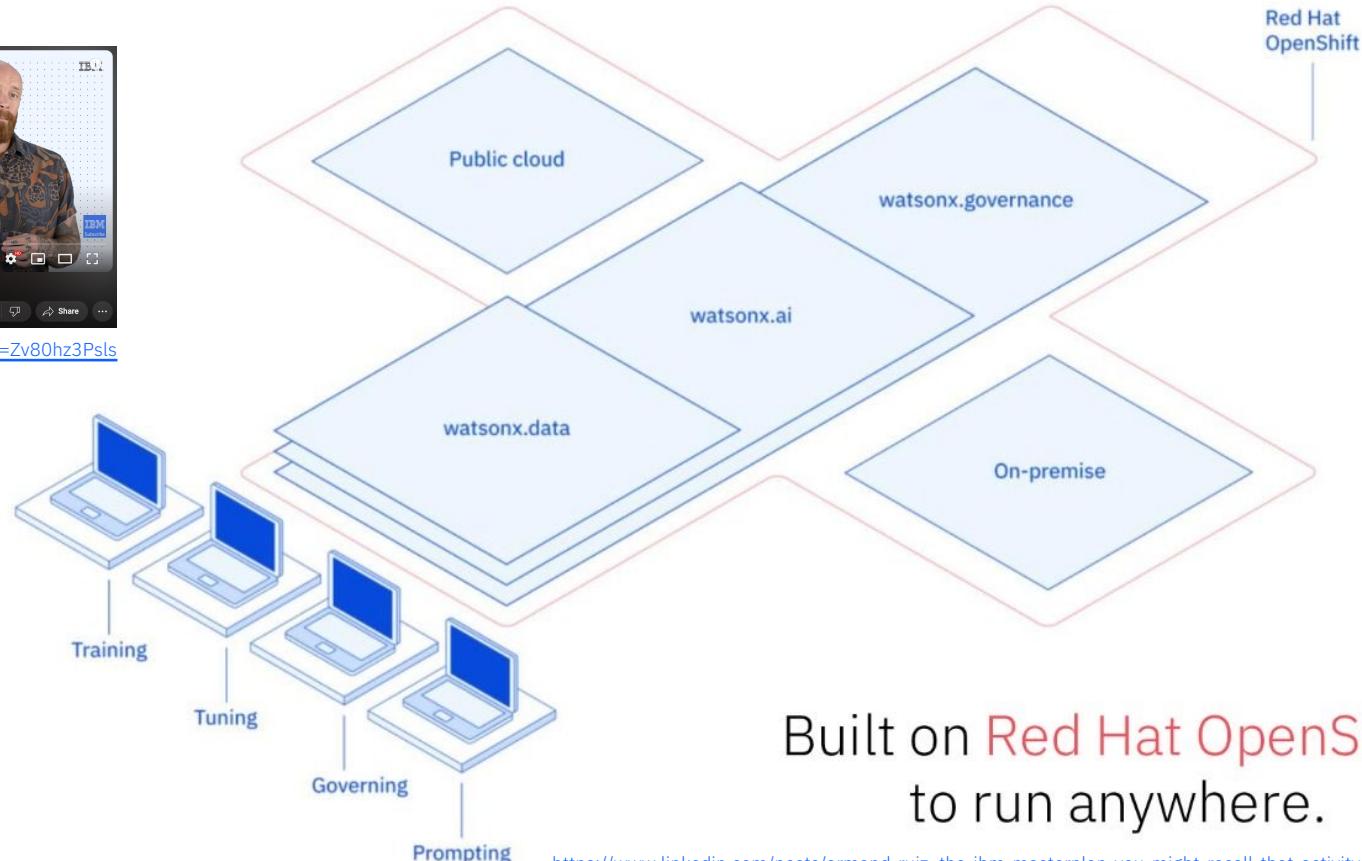
[ibm.com/br-pt/products/watsonx-ai](https://ibm.com/br-pt/products/watsonx-ai)

Put AI to work with **watsonx**.

## Putting AI to work on Hybrid Cloud

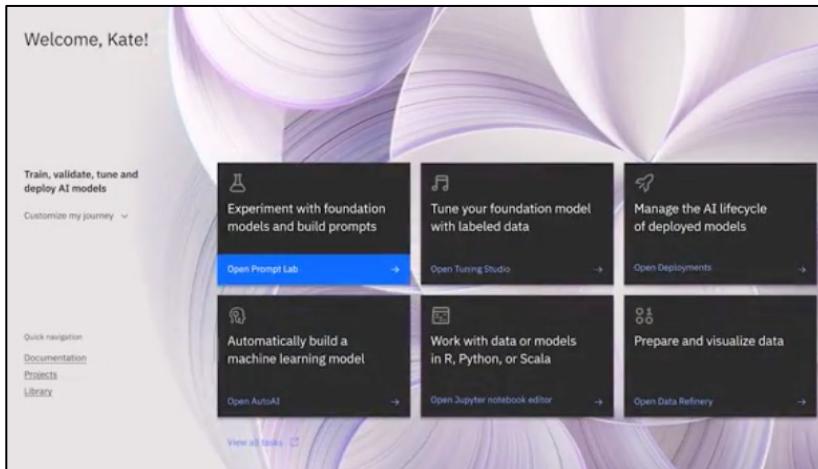


<https://www.youtube.com/watch?v=Zv80hz3PsIs>



## watsonx.ai

Train, validate, tune, and deploy AI models



A next generation enterprise studio for AI builders to train, validate, tune, and deploy generative AI, foundation models, and machine learning capabilities.

The watsonx.ai components include:

- **Foundation Model Library** with IBM and open-source models
- **Prompt Lab** to experiment with foundation models and build prompts for various use cases and tasks
- **Tuning Studio** to tune your foundation models with labeled data
- **Data Science and MLOps** to build machine learning models automatically with model training, development, and visual modeling

# watsonx.ai: Prompt Lab

Experiment with foundation models and build prompts

## Interactive prompt builder

Includes prompt examples for various use cases and tasks

Experiment with different prompts, save and reuse older prompts, use different models and vary different parameters

Experiment with zero-shot, one-shot, or few-shot prompting to get the best results

## Experiment with prompt engineering

Choice of foundation models to use based on task requirements

Prevent the model from generating repeating phrases

Number of min and max new tokens in the response

Stop sequences – specifies sequences whose appearances should stop the model

The screenshot shows the IBM WatsonX Prompt Lab interface. At the top, there's a navigation bar with 'IBM watsonx', a search bar, and various account and location options. The main area is titled 'Prompt Lab' and shows a workspace named 'New (unsaved)'. On the left, a sidebar titled 'Sample prompts' lists several options: 'Summarization', 'Earnings call summary' (selected), 'Meeting transcript summary' (highlighted in blue), 'Scenario classification', 'Sentiment classification', 'Marketing email generation', 'Thank you note generation', 'Named entity extraction', 'Fact extraction', and 'Question answering'. The main content area has tabs for 'Structured' and 'Freeform'. Under 'Set up', there's an 'Instruction (optional)' field with the placeholder 'Write a short summary for the meeting transcripts.' and an 'Examples (optional)' section with a table. The table has two columns: 'Transcript' and 'Summary'. It contains two rows of data. The first row is for 'Meeting transcript summary' and the second for 'Scenario classification'. At the bottom, there's a 'Try' section with a 'Test your prompt' input field, a table for 'Fact extraction' with a single entry '1' and summary 'John Doe 00:00:01.415 --> 00:00:20.675', and a 'Generate' button.

## What IBM offers

# watsonx.ai: Models available

granite.13b 13 billion params decoder only	flan-ul2-20b 20 billion params encoder/decoder	gpt-neox-20b 20 billion params decoder only	mt0-xxl-13b 13 billion params encoder/decoder	flan-t5-xxl-11b 11 billion params encoder/decoder	mpt-instruct2-7b 7 billion params decoder only	llama2 70 billion params decoder only	starcoder 15.5 billion params decoder only
Q&A	Q&A	Q&A	Q&A	Q&A	Q&A	Q&A	CodeGen
Generate	Generate	Generate	Generate	Generate	Generate	Generate	
Extract	Extract		Extract	Summarize		Extract	
Summarize	Summarize		Summarize	Classify		Summarize	
Classify	Classify		Classify			Classify	
<i>IBM Model</i>	<i>Open Source</i>	<i>Open Source</i>	<i>Open Source</i>	<i>Open Source</i>	<i>Open Source</i>	<i>Meta</i>	<i>3rd Party</i>
Why me: Built on enterprise-relevant datasets; IP protections	Why Me: Flexibility	Why Me: Special Characters Context Length	Why Me: Multi-Lingual Model 100+ languages	Why Me: Medium Instruct	Why Me: Small Instruct	Why Me: Chat	Why Me: Code

## What IBM offers

### watsonx.ai: IBM Granite Models

IBM's approach to AI model development is **grounded in core principles of trust and transparency.**

You can use them for:

- Summarization
- Insight extraction and classification
- Retrieval-Augmented Generation (RAG)

These models have been trained on enterprise-relevant datasets across these domains:

- Internet
- Academic
- Code
- Legal
- Finance

Text extraction

**6.48T** of extracted data

Remove duplicates

**4.9T** of deduplicated data

Remove HAP

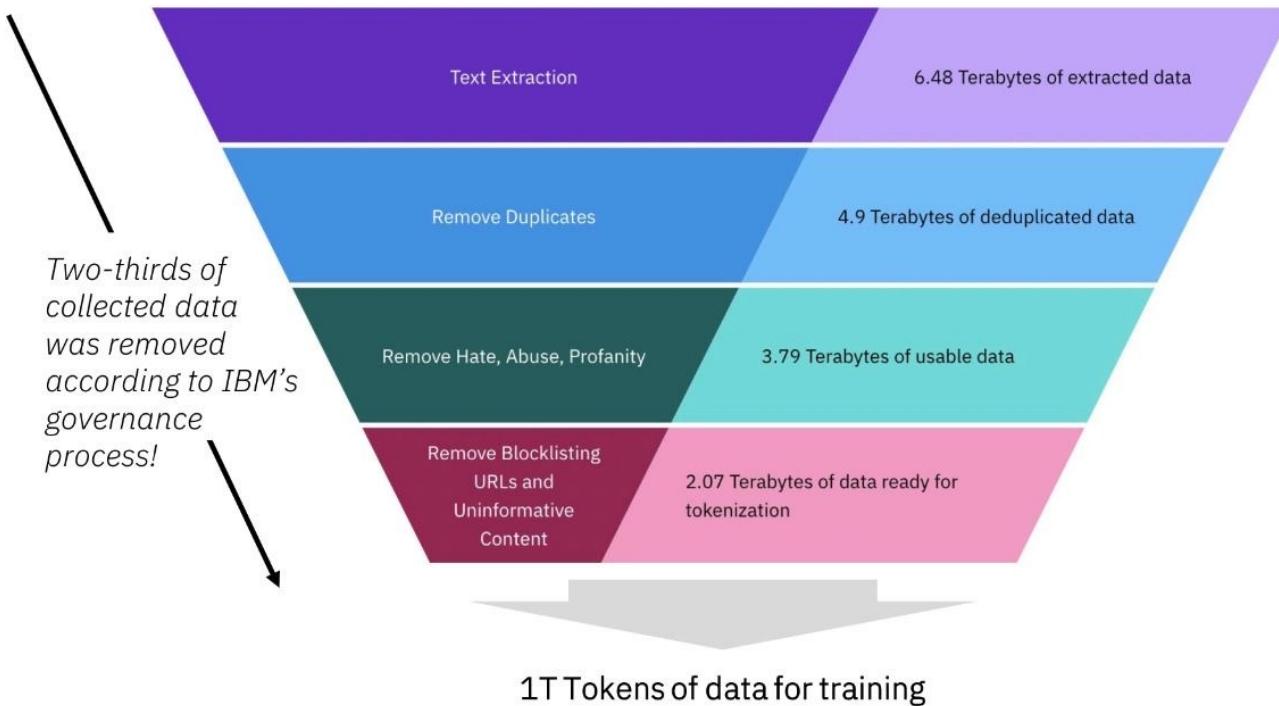
**3.79T** of usable data

Remove poor quality

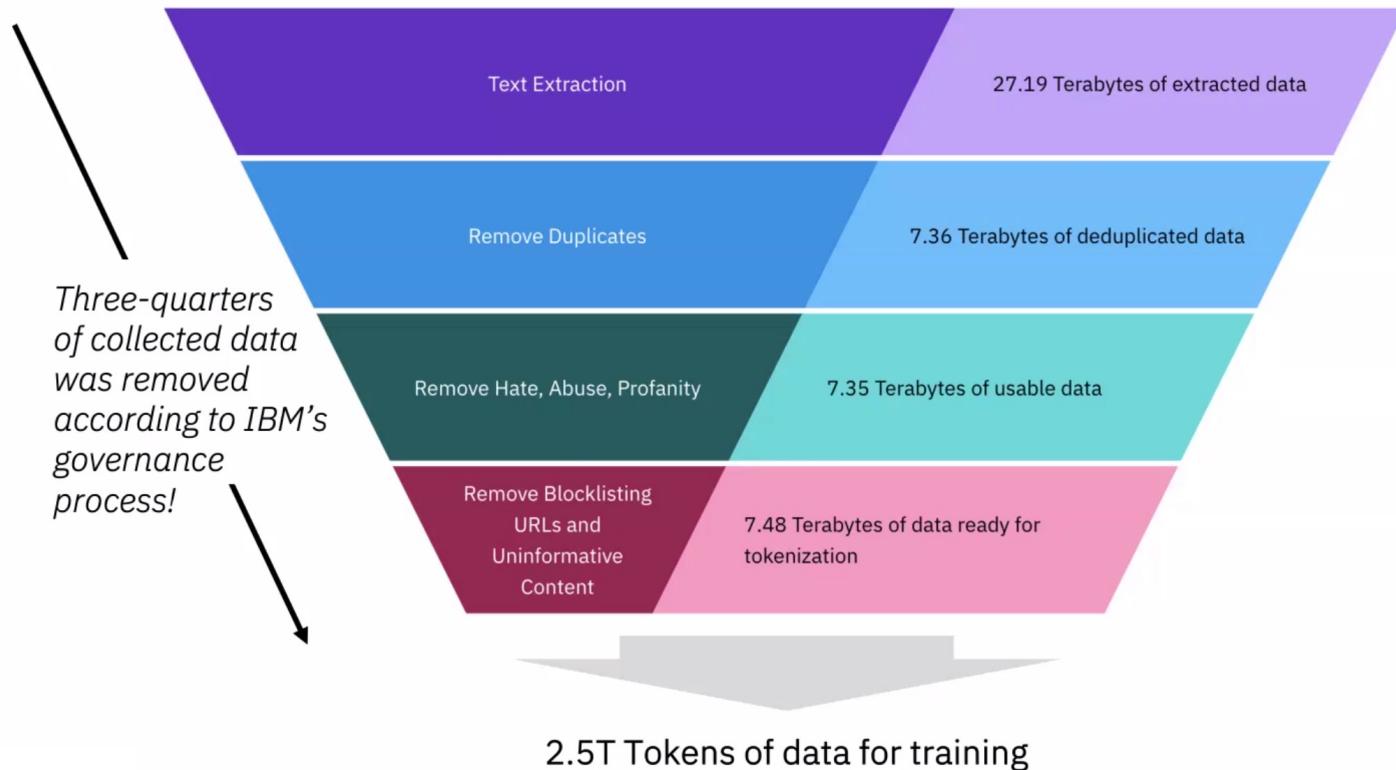
**2.07T** of data for tokenization

## Granite.13b:

### Training data governance funnel



# Granite.13b.v2: Pre-training data governance funnel



## IBM POV: Four core principles to tailor generative AI for enterprise

### Open

- Based on the best AI and cloud technologies available
- Giving access to the innovation of the open community and multiple models

### Targeted

- Designed for targeted business use cases, that unlock new value
- Including curated models that can be tuned to proprietary data and company guidelines

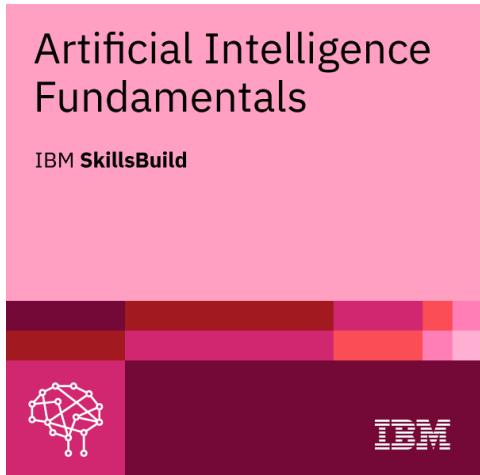
### Trusted

- Built with AI and data governance, transparency, and ethics that support increasing regulatory compliance demands
- Providing guidance on appropriate models to leverage to create real business value with trust

### Empowering

- On a platform to bring your own data and AI models that you tune, train, deploy, and govern
- Running anywhere, designed for scale and widespread adoption to truly create enterprise value

10h  
gratuito em  
português



Plano de aprendizagem

## Fundamentos da inteligência artificial (Ganhe uma credencial!)

Aproximadamente 10 horas 8,222 348

Compartilhar

Em andamento

0%

0 concluídas de 6 obrigatórias

[Atualizar](#)

[Mostrar mais](#)



Cursos

0 de 6 concluído(s)

Optional courses

Opcional

Informações de

credenciais

Opcional

### Sobre este plano de aprendizagem

A inteligência artificial (IA) desperta seu interesse? Aqui está sua chance de se imaginar em uma carreira de IA! Você vai explorar o histórico da IA e descobrir como ela pode mudar o mundo. Ao longo do percurso, você vai descobrir de que maneiras a IA faz previsões, entende linguagem e imagens e aprende usando circuitos inspirados no cérebro humano. Após uma simulação prática na qual você constrói e testa um modelo de aprendizado de máquina, é possível receber dicas de como buscar sua própria carreira na área de inteligência artificial.

Conclua os cursos obrigatórios a seguir para obter uma credencial digital do IBM SkillsBuild reconhecida pelo setor chamada **Fundamentos da inteligência artificial**:

1. Introdução à inteligência artificial
2. Processamento de linguagem natural e Visão por computador
3. Aprendizado de máquina e Deep learning
4. Execute modelos de IA com o IBM Watson Studio
5. Ética da IA
6. Seu futuro na IA: o panorama de trabalho

[Cancelar matrícula](#)

Ações

<https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030>



# Prompt Engineering with watsonx.ai

Welcome! The goals of this workshop are the following:

- Gain a comprehensive understanding of prompt engineering
- Learn techniques to achieve the best results with Large Language Models
- Apply learnings through completion of a diverse set of prompt engineering exercises

<https://ibm.biz/watsonx-prompt>  
Code: watsonxlab

## About this workshop

Part art, part science, prompt engineering is the process of crafting input text to fine-tune a given large language model for best effect.

Foundation models have billions of parameters and are trained on terabytes of data to perform a variety of tasks, including text-, code-, or image generation, classification, conversation, and more. A subset known as large language models are used for text- and code-related tasks.

When it comes to prompting these models, there isn't just one right answer. There are multiple ways to prompt them for a successful result.

## Explore Content

All Courses 129 Guided Projects 284

Sort ▾

Reset Filters

Results (85)

- Skills
- Tier
- Topic
- Difficulty
- Platform
- Language
- Offered By
- Industries
- Ratings



Beginner Course

Prompt Engineering for Everyone

Prompt Engineering, Generative AI, LLM, +1 more

9.72k+ Enrolled

5 Hours

★★★★★ (1.03k+)



Beginner Course

Introducing AI

Artificial Intelligence

6.34k+ Enrolled

1 Hour

★★★★★ (1.65k+)



Beginner Course

AI Ethics

Artificial Intelligence, Ethics, Bias, +2 more

2.29k+ Enrolled

1 Hour

★★★★★ (435)



Beginner Course

AI Concepts

Artificial Intelligence, Machine Learning, Deep...

1.73k+ Enrolled

1 Hour

★★★★★ (348)



Intermediate Guided Project

Build a Chatbot to Analyze PDF Documents Using LLM

Python, Generative AI, LLM, +2 more

570 Enrolled

60 Min

★★★★★ (58)



Beginner Guided Project

Build a Chatbot in less than an hour with watsonx!

Chatbots, Generative AI, Artificial Intelligenc...

600 Enrolled

1 Hour

★★★★★ (103)



# Programa de cursos integrados Generative AI Fundamentals

Unlock and leverage the potential of generative AI. Learn how you can use the capabilities of generative AI to enhance your work and daily life.

Ministrado em Inglês



Instrutores: [IBM Skills Network Team](#) + Mais 2

Cadastre-se gratuitamente  
Inicia em Nov 30

Auxílio financeiro disponível

# Saiba mais...

Desbloqueando a produtividade com watsonx.ai: Como a IA pode revolucionar seu negócio?

Alan Braz  
TDC Future 8/12/2023

WatsonX AI logo

Desbloqueando a produtividade com watsonx.ai:Como a IA pode revolucionar seu negócio?  
Não listado

Pull reCast - Desenvolvendo sua Eminência...  
1,11 mil inscritos

Analytics Editar vídeo

Like 2 Compartilhar Promover Download ...

<https://youtu.be/UVMcgWUohwo>



[Hipsters: Fora de Controle #42](#)

Acompanhe as notícias e  
mais eventos no [@pullrecast](#)

2:10

pullrecast

70 posts 315 followers 27 following

Pull reCast | Podcast sobre tecnologia  
Digital creator  
Explorando o universo tech!  
Apresentado por @alanbraz e @matbitit  
Não representa opiniões da IBM... more  
See Translation  
pullrecast.dev and 2 more

Followed by scriptorebraz, matbitit and 87 others

Following Message

YouTube 1 IBM Next C... Alan Matheus Pull reCast

Conheça o Pull reCast!

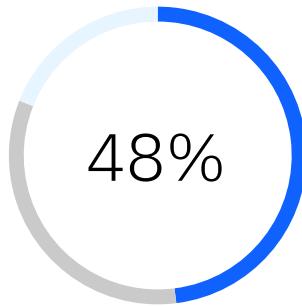
WatsonX SEVILLA FC



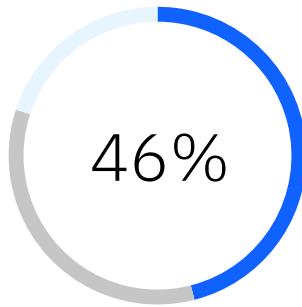
# Generative AI adoption considerations, inhibitors and fears

80% of business leaders see at least one of these ethical issues as a major concern

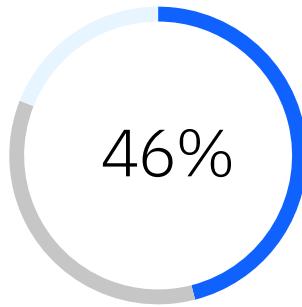
Explainability



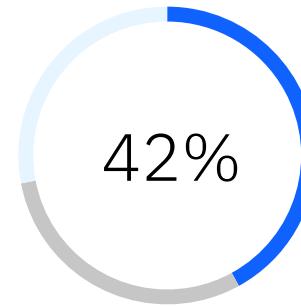
Ethics



Bias



Trust



Believe decisions made by generative AI are not sufficiently **explainable**.

Concerned about the safety and **ethical** aspects of generative AI.

Believe that generative AI will propagate established **biases**.

Believe generative AI cannot be **trusted**.

■ Agree ■ Neutral ■ Disagree

Enterprises need more than an AI solution - they need a comprehensive and sound strategy for generative AI.