INSTITUTO DE PESQUISAS TECNOLÓGICAS

COMUNICAÇÃO TÉCNICA

Nº 180013

Como imagens e texto conversam: explicando vision-language models

Gustavo Torres Custódio
Adriana Camargo Brito

Palestra apresentada no Start IA, SEBRAE. 80 slides.

A série "Comunicação Técnica" compreende trabalhos elaborados por técnicos do IPT, apresentados em eventos, publicados em revistas especializadas ou quando seu conteúdo apresentar relevância pública.
PROIBIDO A REPRODUÇÃO, APENAS PARA CONSULTA.

Instituto de Pesquisas Tecnológicas do Estado de São Paulo S/A - IPT
Av. Prof. Almeida Prado, 532 | Cidade Universitária ou
Caixa Postal 0141 | CEP 01064-970
São Paulo | SP | Brasil | CEP 05508-901
Tel 11 3767 4374/4000 | Fax 11 3767-4099

www.ipt.br

start IA

Kick-Off

ciclo 2

Sebrae for Startups

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Cronograma

09h00 – **Boas vindas**

09h30 – **Oportunidades Sebrae for Startups & IPT**

10h00 – **Sobre os programas**

10h30 – **Benefícios Abstartups**

11h00 – **Palestra IA**

12h30 – **Encerramento**

Realização: Sebrae for Startups

Execução: IPT OPEN | ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Instituto de Pesquisas Tecnológicas do Estado de São Paulo

Marcus Leite
Coordenador - Economia Criativa e Startups

Alex Fedozzi Vallone
Gerente de Programas, Inovação e IPT Open

Realização:
Sebrae for Startups

Execução:
IPT OPEN
ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

SEBRAE SP

Maria Augusta Pimentel Miglino
Gestora estadual de programas -
Sebrae For Startups

Realização:
Sebrae for Startups

Execução:
IPT OPEN
ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

Sebrae
for Startups

Sebrae
r Start

Olá,

Somos o Sebrae For Startu

programa do Sebrae São Paulo

RESULTADOS 21-2024

EMPREENDEDORES NO CENTRO DE TUDO

3.000 Inscritos
Top 30 – 6 de São Paulo
Top 10 – 3 de São Paulo
Top 01 – DIO (São Paulo)

2X
ACELERADORA

1X
AGENTE DE FOMENTO

1X
ED. EMPREENDEDORA

CIÊNCIA E TECNOLOGIA PARA MUDAR O MUNDO

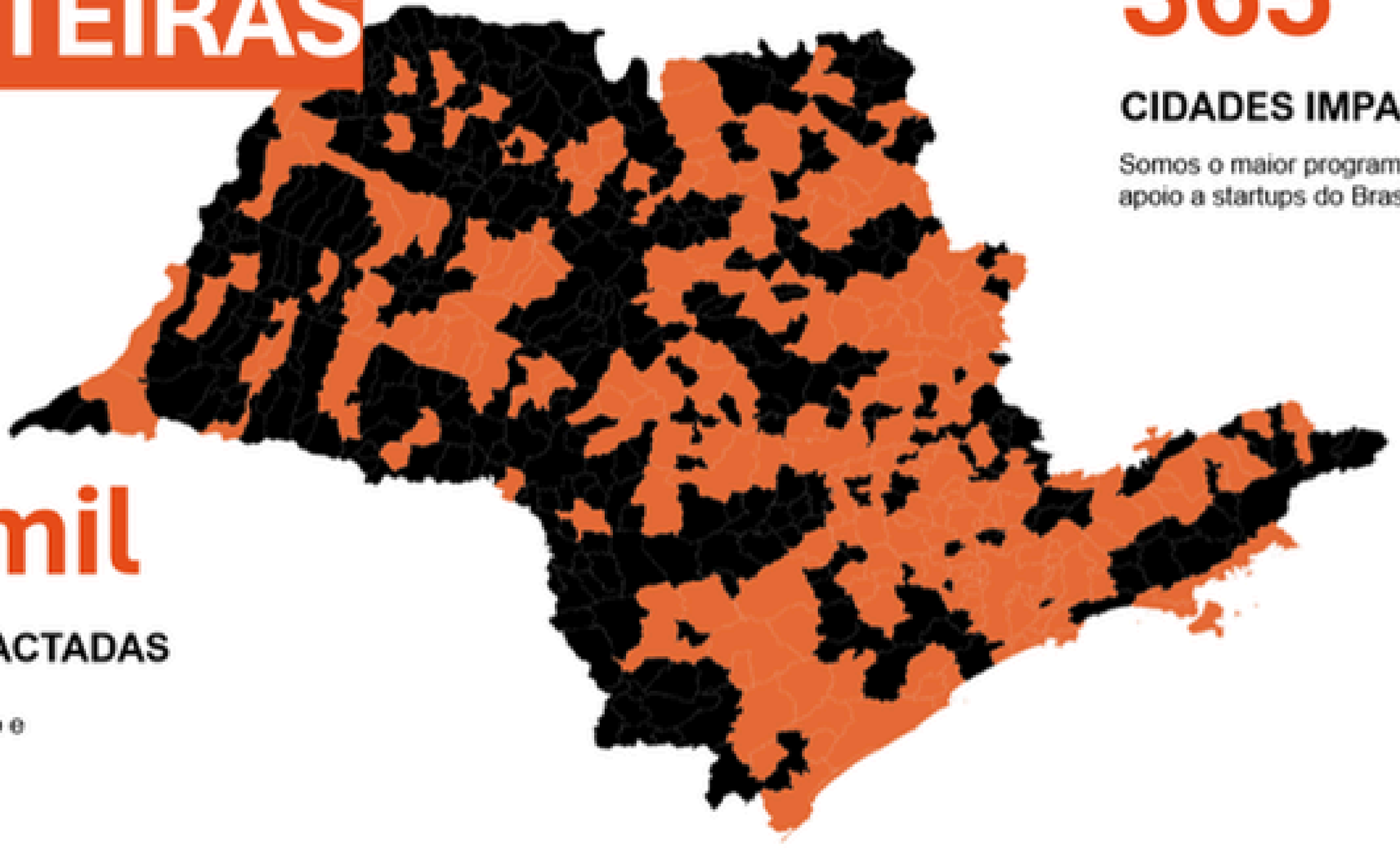
+4 mil
STARTUPS ATENDIDAS

+200
FACULDADES ENGAJADAS

+90 mi
SUBVENÇÃO FAPESP

+40 mi
PARTICIPAÇÕES

INOVAÇÃO SEM FRONTEIRAS



365

CIDADES IMPACTADAS

Somos o maior programa de apoio a startups do Brasil

+20 mil

PESSOAS IMPACTADAS

Ações de sensibilização e programas

NÃO FAZEMOS NADA SOZINHOS

+60

PARCEIROS

+600

MENTORES

PROGRAMAS

spark
start
speed
GINGA
iMERSO
GINGA
prototipa
GINGA
ideias
Curiosidade
Ideação
Validação
Tração
Crescimento

Subvenção (FAPESP)

LINHAS DE SUBVENÇÃO

CONVÊNIO FAPESP

R$ 150 M

R$ 75 M (SEBRAE) + R$ 75 M (FAPESP)

Acesso ao mercado **Desenvolvimento**

AÇÕES

- Lançamento de Editais
- Programa de mentorias
- Apoio para escrita de projeto

LINHAS DE SUBVENÇÃO
CONVÊNIO FAPESP

R$ +90MI

Comprometidos

12 Editais lançados

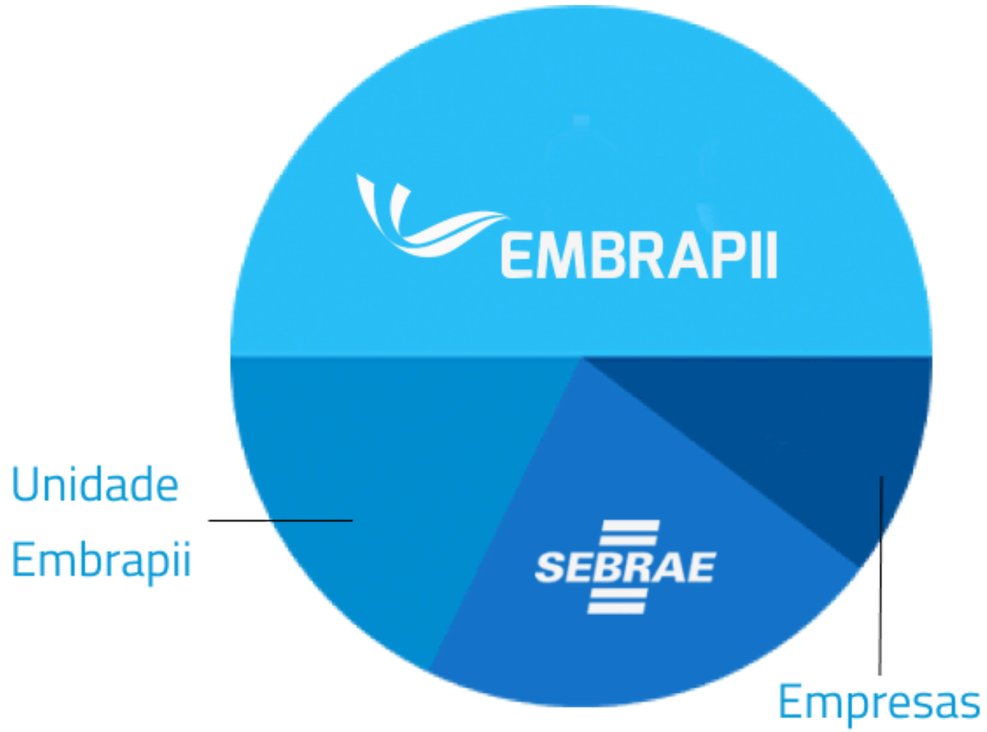
176 empresas apoiadas

Subvenção (EMBRAPII)

COMO FUNCIONA?

A EMBRAPII aporta até 50% do montante total do projeto e o pagamento do valor restante é negociado entre a Startup/MPE e Unidade Embrapii.

O Sebrae cobre até **70% da contrapartida estabelecida**



start

deeptech

start | Sebrae for Startups

PRÉ-ACELERAÇÃO

start digital

ACOMPANHAMENTO & MENTORIA

MEETUPS & CAPACITAÇÃO

start deeptech

ACOMPANHAMENTO & MENTORIA

MEETUPS & CAPACITAÇÃO

CONSULTORIA DE INOVAÇÃO

SERVIÇOS TECNOLÓGICOS

start bio

DIAG. MATURIDADE TECNOLÓGICA

MENTORIA TECNOLÓGICA

SERVIÇOS LABORATORIAIS

start ia

DIAG. MATURIDADE TECNOLÓGICA

MENTORIA TECNOLÓGICA

start

Helmo Oliveira
Start Digital

Maria Augusta Miglino
Start Deeptech, Start Bio,
Start IA

Catarina Cano
Fapesp

Eduardo Gomide
Incubadoras e
Parques Tecnológicos

Bruno Brigida
Ginga Afrotech Hub

Larissa Vieira
Start Deeptech, Start Bio,
Start IA - (by IPT)

Ricardo Bueno
Start Deeptech, Start Bio,
Start IA - (by IPT)

Tatyane Sales
Start Mercado
(by IPT)

Alexandre Rodrigues
Start Digital
(by ABStartups)

Para mais oportunidades , scaneie o QR code

start IA

Instituto de Pesquisas Tecnológicas do Estado de São Paulo

Alex Fedozzi Vallone

Gerente de Programas, Inovação e IPT Open

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

Watch video on YouTube

Error 153

Video player configuration error

start IA

O que somos?

Empresa pública estadual
Vinculada a SCTI
Fundada em 1899

126 anos de contribuições para a sociedade

+ de 1000 funcionários e colaboradores

50% da receita em inovação

O que fazemos?

Soluções tecnológicas para a indústria, governo e sociedade

+ de 3170 clientes atendidos

+ de 16200 documentos técnicos emitidos

+ de 2000 procedimentos e ensaios no portifólio

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Tecnologias Digitais

Bionano - manufatura

Energia

Habitação e edificações

Materiais avançados

Ensino Tecnológico

Tecnologias Regulatórias e Metro - lógicas

Cidades infra-estrutura e meio ambiente

Unidades de Negócios

Realização:

Sebrae for Startups

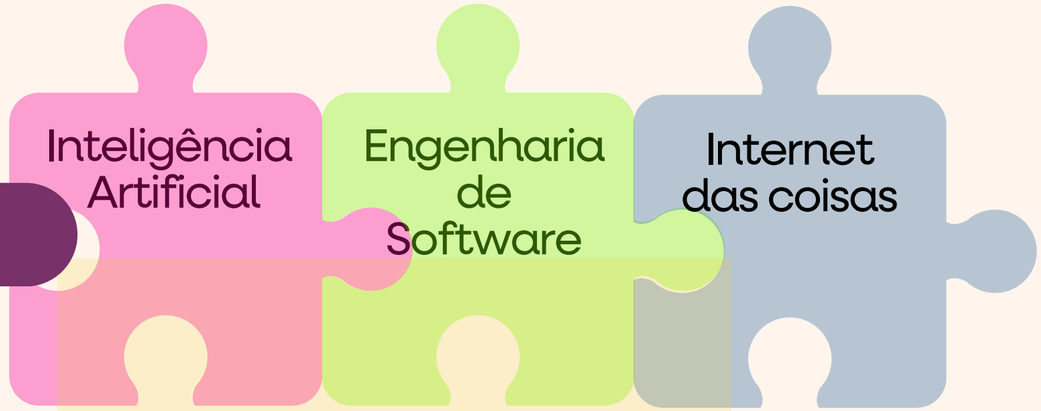
Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Tecnologias Digitais



Cidades inteligentes

- Análise de dados
- Apps para gestão
- Sensores inteligentes
- Salas de Situação
- Conectividade

Indústria 4.0

- Interoperabilidade de sistemas e dispositivos
- Conformidade e testes
- IA para a indústria
- Realidade virtual e aumentada

Mobilidade e Transporte

- Transportes inteligentes
- Big data e Analytics
- Simulações
- Verificações em aeroportos, rodovias, etc.

Inovação em TIC

- Sistemas operacionais modernos
- Acessibilidade
- Cybersegurança
- Bioinformática

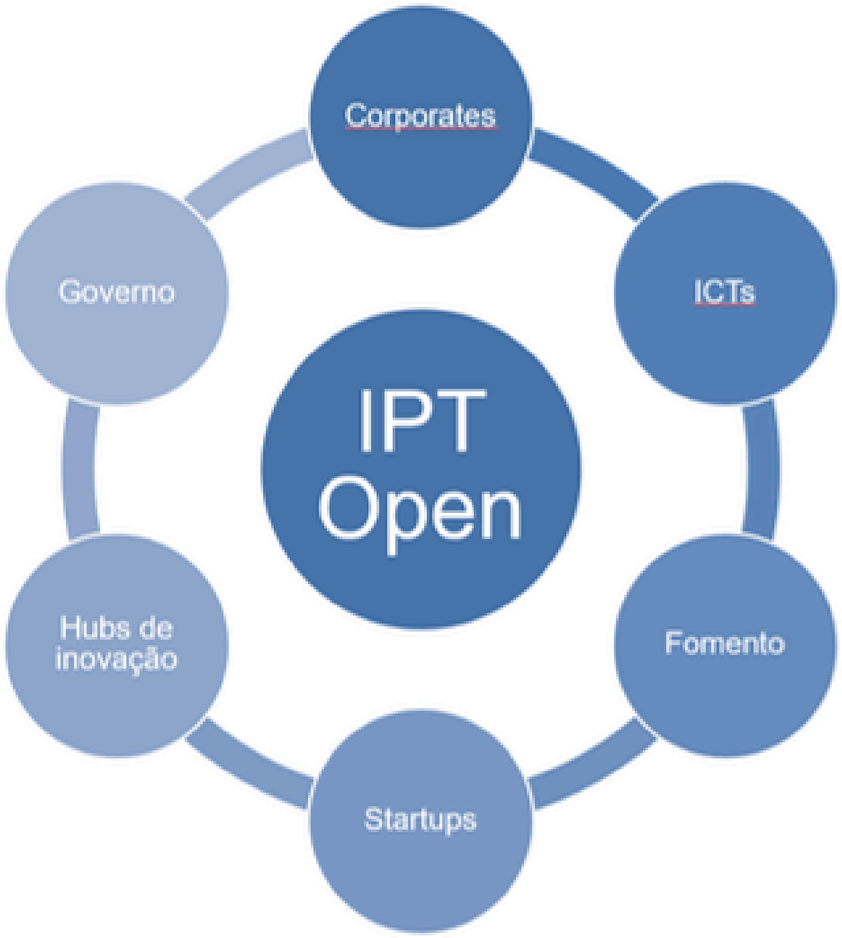
Realização: Sebrae for Startups

Execução: IPT OPEN — ipt Instituto de Pesquisas Tecnológicas

start IA

Ecossistema IPT

Conectamos pessoas, tecnologias e organizações para **transformar conhecimento em impacto**.



Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

O que fazemos?

IPT Open é o orquestrador do ecossistema de inovação IPT

Realização: Sebrae for Startups

Execução: IPT OPEN | ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

Conheça mais de cada área:

start IA

Programas Desenvolvimento Tecnológico Deeptechs

Luciano Avallone Bueno
Pesquisador Líder Deeptech - IPT Open

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Programas IPT – SEBRAE (2022 – 2026)

Realização: Sebrae for Startups

Execução: IPT OPEN | ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

NOSSOS NÚMEROS

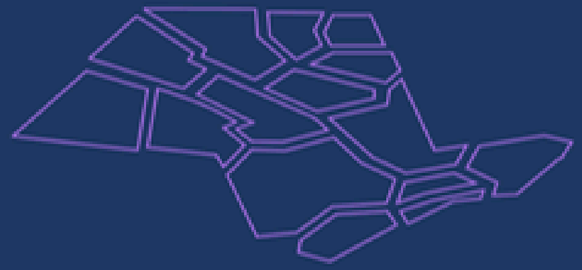
the web Frame

3646 STARTUPS AVALIADAS - 22/25

STARTUPS ATENDIDAS **364**

57 MUNICÍPIOS DE SP COM STARTUPS ATENDIDAS

+20 LABORATÓRIOS



PESQUISADORES APOIADOS **+600**

423 INSTITUIÇÕES NOS PROGRAMAS

DIAGNÓSTICOS EXECUTADOS **+400**

+30k HORAS DE MENTORIAS

EQUIPAMENTOS **+ 70**

08 ÁREAS ENVOLVIDAS

Realização: Sebrae for Startups

Execução: IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Sobre o Programa

Adriana Camargo de Brito
Pesquisadora IPT-SIAA

Realização:
Sebrae for Startups

Execução:
IPT OPEN
ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Tópicos abordados

Aprendizado de Máquina
Aprendizado Profundo
Visão Computacional
Processamento de Linguagem Natural
Análise de dados

Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Foco das mentorias

Mentor: conselheiro, instrutor
Recomendações
Sugestões
Exemplos

Realização: Sebrae for Startups

Execução: IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Importante

Participação da equipe técnica
Ter tempo para realização de tarefas
Avisar com 24 h de antecedência se precisar faltar em reunião
Estar sempre em contato com seus mentores

start IA

Entregáveis

Plano de Trabalho
Relatório Final

Realização: Sebrae for Startups

Execução: IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Cronograma

🛠️ **Ajustes Iniciais**
Reuniões online de ajuste do Plano de trabalho: 10/11 a 10/12/2025
📢 **Discussões DMT**
Apresentação e Devolutivas DMT: 24/11 a 10/12/2025
📄 **Entrega preliminar**
Plano de Trabalho: 03/12/2025
🎓 **Mentorias (reuniões quinzenais)**
08/12/2025 a 24/04/2026
📝 **Entrega final**
Preenchimento do DMT final e Devolutivas: 13/04 a 30/04/2026
Entrega do Relatório Final: 27/04 a 06/05/2026
Encerramento: 07/05/2026

Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start

Mentores

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start

Mentores

	Mentor principal	Mentor
Data Overseas	Pacheco	Aline
Organize meu Condomínio	Cassia	Eric
Syncorp Tecnologia	Cassia	Silles
Remy.AI	Cassia	Gustavo
Marelo	Pacheco	Angelina
Imans	Thiago	Luciana
Kiskadee Analítica	Thiago	Pakito
CompraMais App	Cassia	Adriana
TrFlow	Thiago	Pakito
Inova Cartórios	Pacheco	Adriana

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Palestras mensais

- **Nov/25** Agentes de IA: Sistemas com Licença para Automatizar
- **Dez/ 25** Como fazer uma pesquisa cientifica aplicada?
- **Jan/ 26** Detecção de Objetos com YOLO e Transformers: Uma Nova Era
- **Fev/26** IA e Smart Cities
- **Mar/26** Avanços Recentes em Reconhecimento Óptico de Caracteres (OCR)

Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Benefício mentorias Abstartups

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

Somos a **Associação Brasileira de Startups**, uma organização sem fins lucrativos feita por startups, para startups.

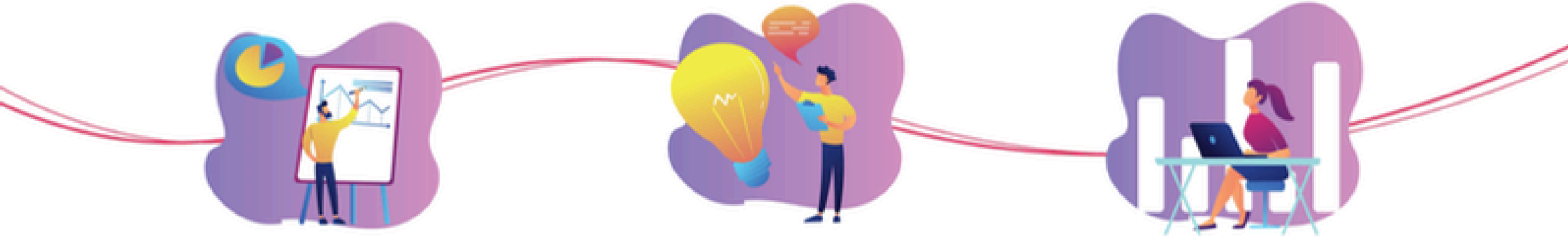
Também (ou mais) conhecida como Associação Brasileira de Startups, desde 2011 existindo graças, e para, as startups.

Nós inspiramos, capacitamos, conectamos e advogamos pelas startups, porque acreditamos no papel da inovação como motor de transformação positiva para o país.

ABSTARTUPS
ASSOCIAÇÃO BRASILEIRA DE STARTUPS

ATUAÇÃO DA ABSTARTUPS

Atuamos no ecossistema por meio de três pilares:

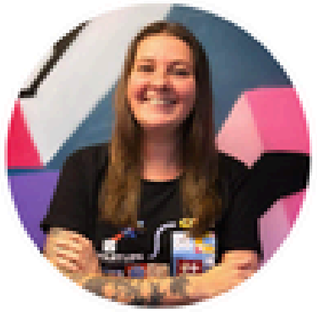


INFORMAÇÃO

PROMOÇÃO

REPRESENTAÇÃO

NOSSO TIME!



Cláudia Schulz
CEO
Gestão
Administração

André Rod
Coordenador de Comunidades
Programas Especiais
Programas

Victória Zanini
Analista de Comunidades
Mentor Especialista
Guias de Acompanhamento

Anthony Alves
Analista de Comunidades
Mentor Especialista

Elisangela Firmino
Analista de Comunidades
Mentor Especialista

Camila Andrade
Trainee de Comunidades
Mentor Especialista

start

RESPONSABILIDADES DA ABSTARTUPS

ATENDIMENTO

REALIZAÇÃO DE ATENDIMENTOS ONLINE COM STARTUPS;

DESAFIOS

ENTENDIMENTO DOS DESAFIOS DA STARTUP;

MENTORIAS

CONEXÃO COM MENTORES;

DIRECIONAMENTO

OBSERVAR E DIRECIONAR STARTUPS

EVOLUÇÃO

DIAGNOSTICAR A EVOLUÇÃO DA STARTUP

Sebrae for Startups

ABSTARTUPS

Quem está com a Startup nesta jornada?

Existem duas pessoas que irão acompanhar a Startup nesta jornada de desenvolvimento.

start

Guia de Acompanhamento

Responsável por realizar as sessões de acompanhamento e diagnosticar, junto à Startup os temas de suas mentorias a partir da trilha destinada!

Mentores especialistas

Responsável por orientar a startup, indicando os melhores caminhos e as melhores formas de superar os desafios, por meio de metodologias e ferramentas.

Sebrae for Startups

ABSTARTUPS

ATENÇÃO!

MELHOR E-MAIL!
OLHE O SPAM!

os contatos são realizados via e-mail, mantenham-se atentos ao e-mail cadastrado!

CONQUISTAS

Aplique e compartilhe o aprendizado!

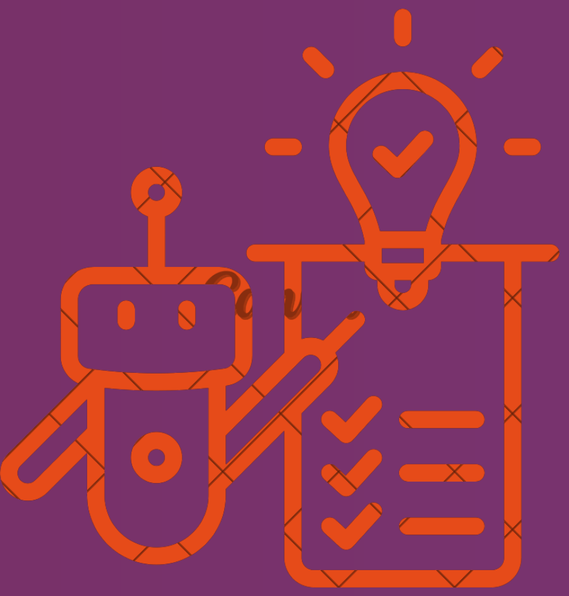
SEJA UM BOM OUVINTE!

COMUNICAÇÃO!

Seja pró-ativo e contate a Abstartups ou mentores quando necessário!

start IA

COMO IMAGENS E TEXTO CONVERSAM: EXPLICANDO VISION-LANGUAGE MODELS



com **Gustavo Torres Custodio**

Apoio: Adriana Camargo de Brito
Mentores de IA do programa no IPT-SIAA

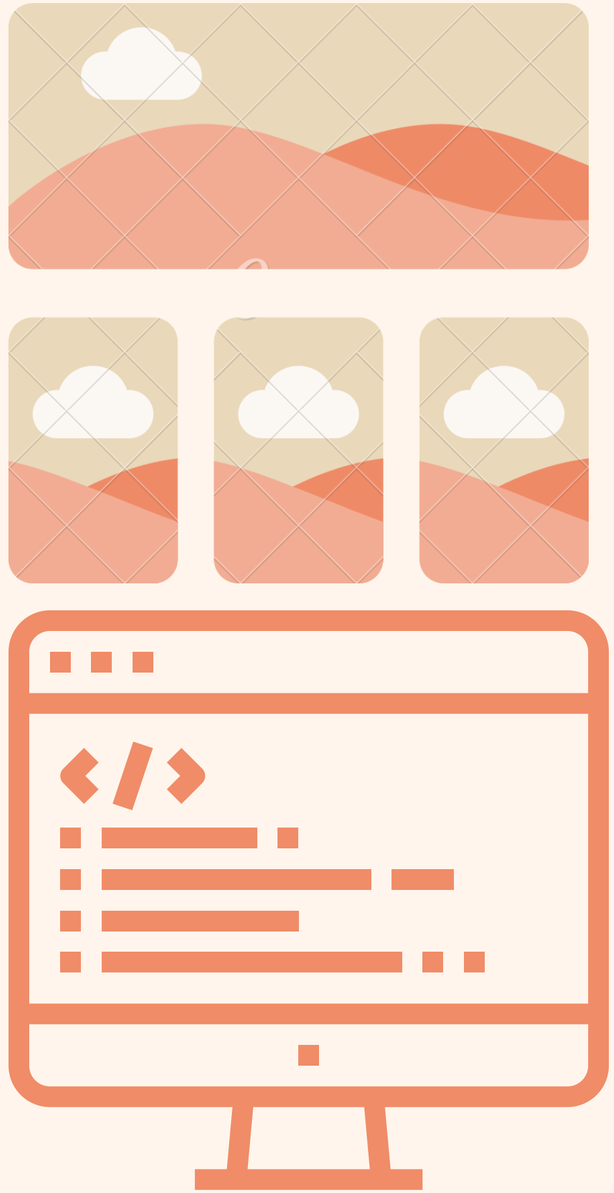
Realização: Sebrae for Startups

Execução: IPT OPEN · ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Introdução

- Legendas transmitem o significado de uma imagem
- Proporcionam entendimento contextual
- Mas, criar legendas manualmente para muitas imagens:
  - Demora
  - É pouco escalável
- Necessidade: gerar legendas automaticamente.



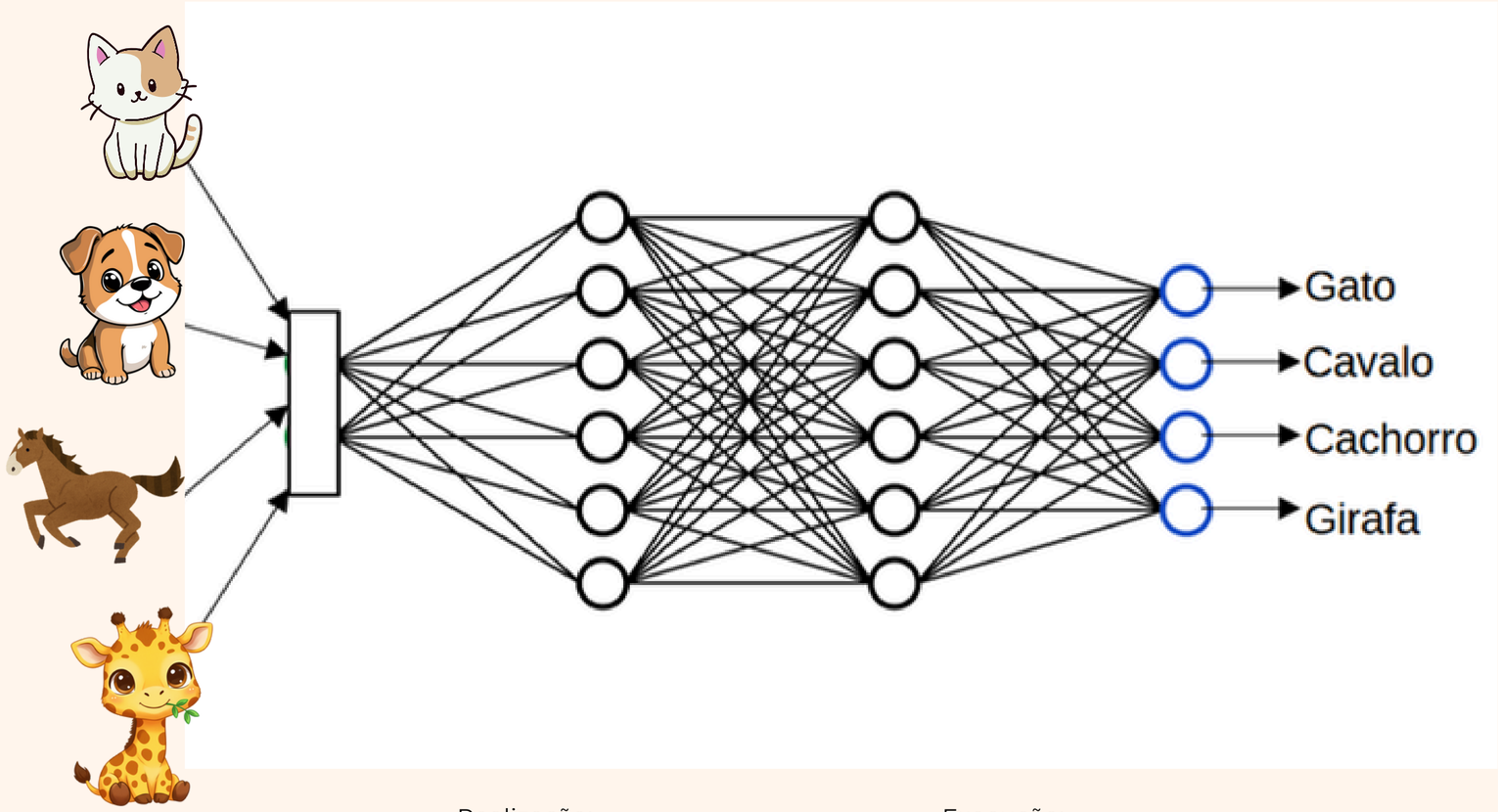
Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Introdução

- Legendas automáticas para imagens
- Pode-se usar rede neural
- Classificação.



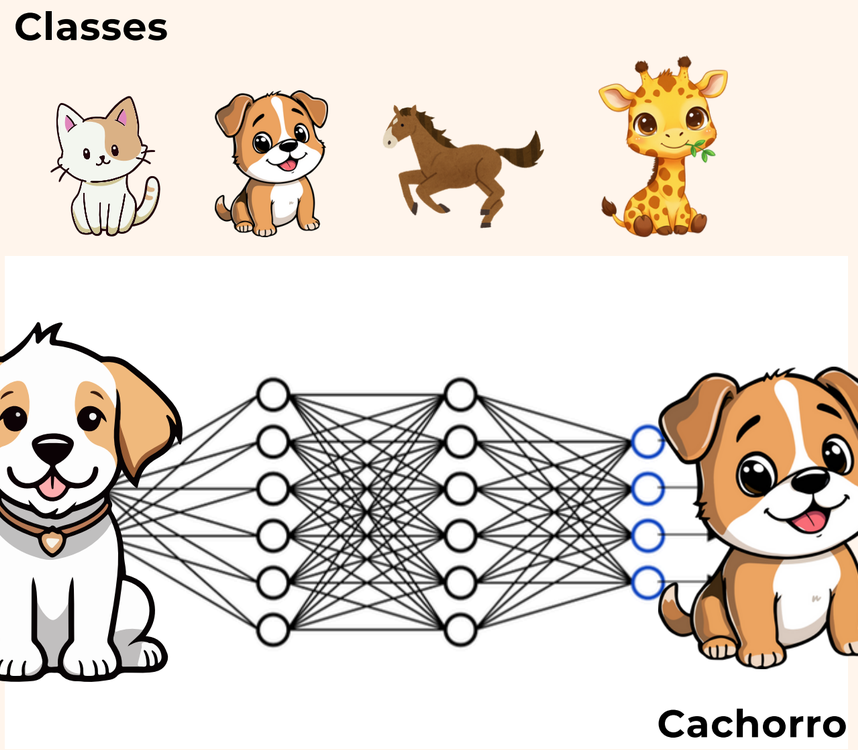
Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Introdução

- Cada animal representa uma classe
- Treino: a rede recebe múltiplas imagens e rótulos indicando qual é o animal de cada imagem
- Com o treinamento a rede consegue "prever" a classe do animal em novas imagens

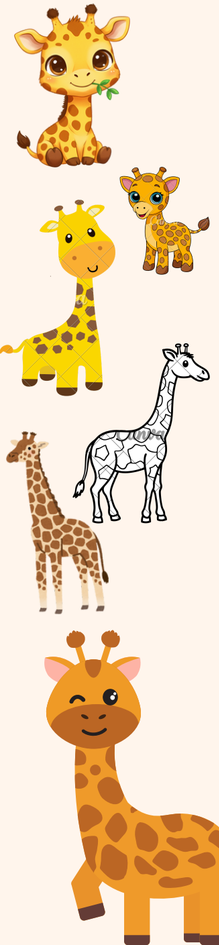


Realização: Sebrae for Startups
Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

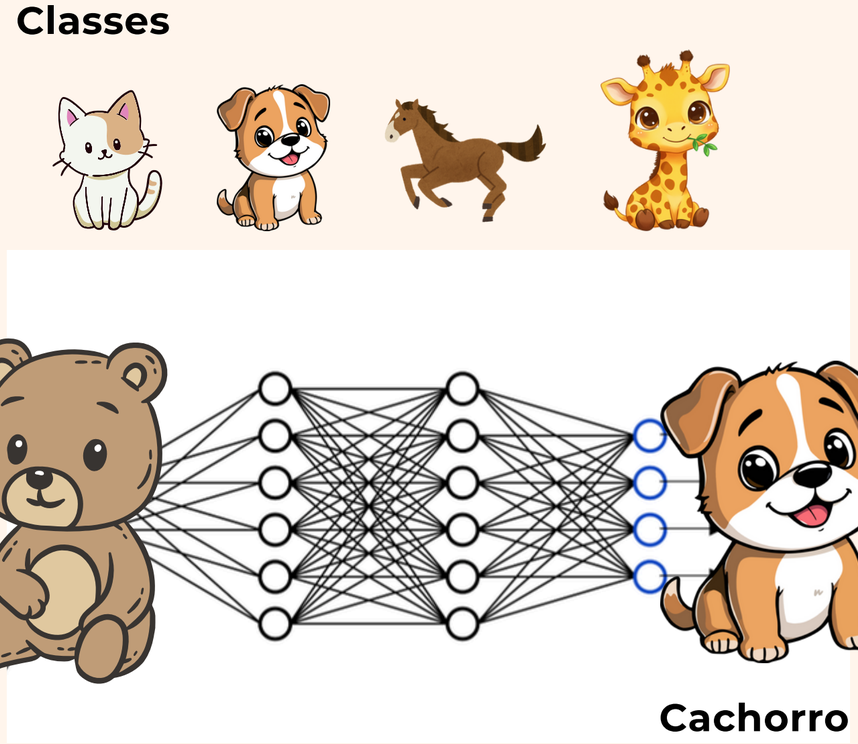
start IA

Introdução

Problemas



- Para treinar: precisa de muitas imagens diferentes de cada tipo
- Necessário treinar a rede do zero
- A rede só prevê uma das quatro classes mostradas no treinamento
- Mesmo que o objeto a ser classificado não seja de nenhuma dessas classes
- **A rede neural não compreende relações entre texto e imagens**



Realização: Sebrae for Startups

Execução: IPT OPEN, ipt Instituto de Pesquisas Tecnológicas

start IA

VLMs

- Modelos multimodais: modelos de IA que integram informações de fontes diferentes (imagem-texto, áudio-texto, etc).
- VLMs (Vision-Language Models): modelos que entendem imagens e texto
- VLMs são possíveis graças à combinação de:
  - modelos de visão computacional (que interpretam imagens)
  - modelos de linguagem (que entendem e geram texto).

Realização: Sebrae for Startups
Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

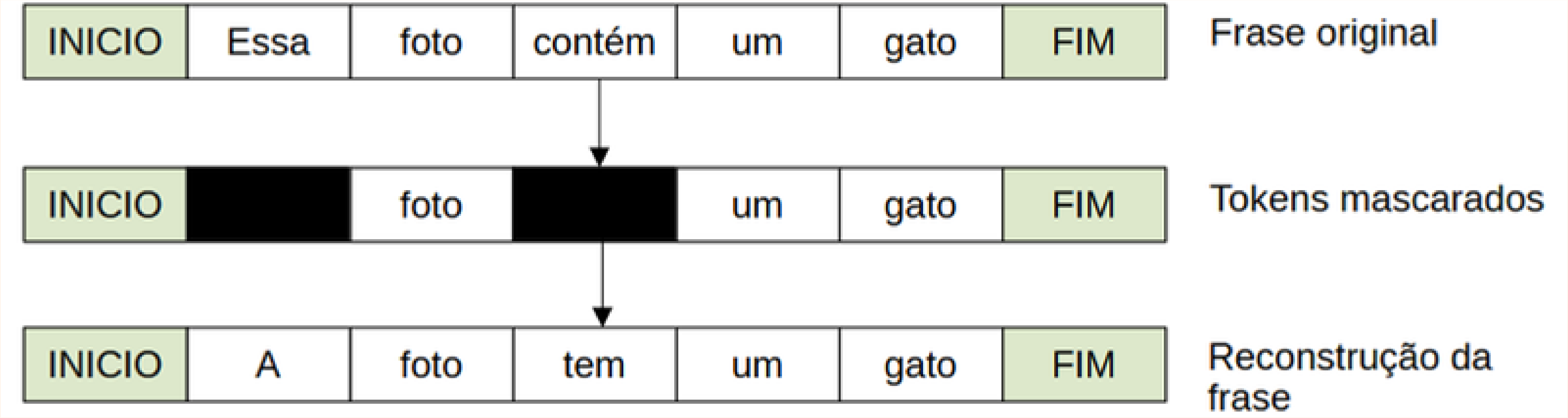
VLMs

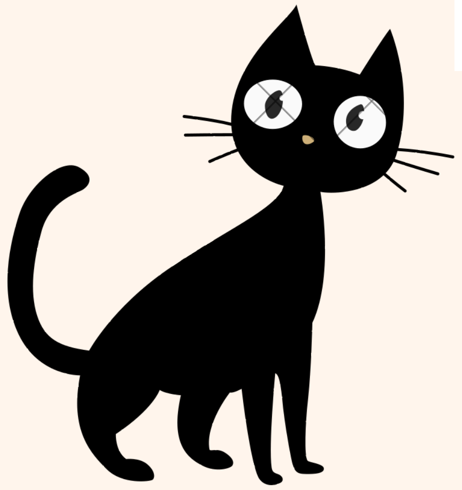
- Modelos de visão e linguagem modernos são muito similares em como aprendem e representam informações
- Isso permite que muitas ideias de processamento de linguagem natural sejam utilizadas em visão computacional e vice-versa

start IA

VLMs

- Treinamento por MLM (Masked-Language Modeling) em modelos de linguagem.





Realização:
Sebrae for Startups
Execução:
IPT OPEN
ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

VLMs

- Treinamento por MAE (Masked AutoEncoder) em modelos de visão.

Realização:

Execução:

start IA

VLMs

- Modelos de linguagem e de visão representam informações por embeddings (sequência de números longa com tamanho fixo)
- Também chamamos isso de espaço latente.
- O espaço latente não diz muita coisa para nós, mas diz muito para os modelos de IA.

Um gato deitado no sofá olhando para frente.

0.4	0.1	0.3	...	0.2

0.7	0.3	0.6	...	0.9

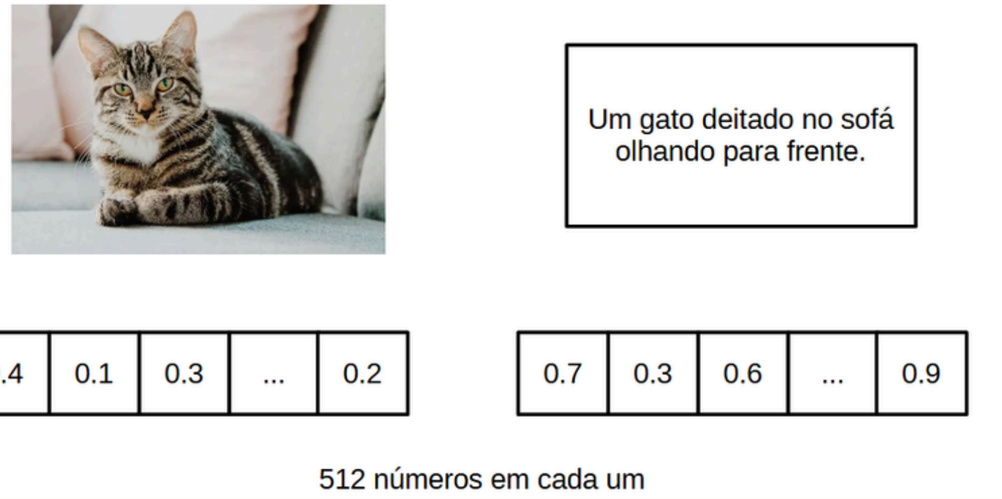
Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

VLMs

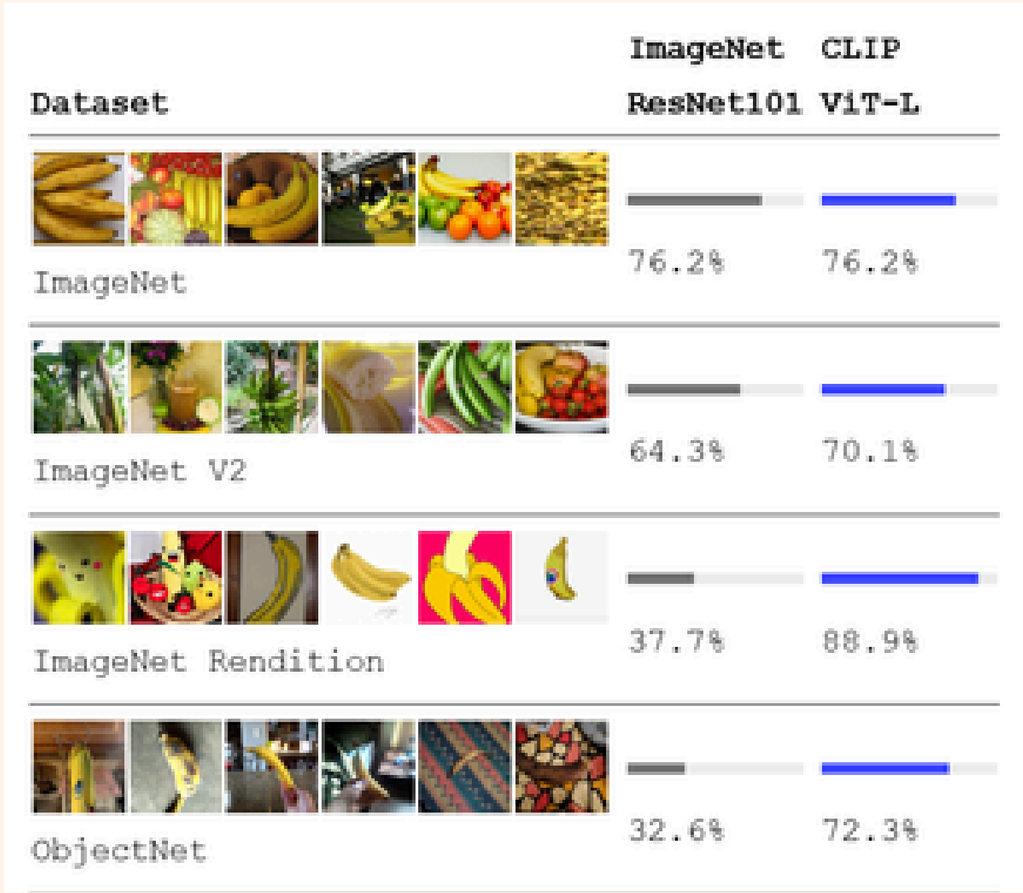
- Agora precisamos fazer esses embeddings conversarem
- conectar os embeddings de imagem com os embeddings da legenda correspondente.
- Um VLM como o CLIP (Contrastive Language-Image Pretraining) faz essa conexão.



start IA

VLMs

- O CLIP foi um dos primeiros modelos a unir visão e linguagem de forma efi ciente
- O CLIP é um modelo disponibilizado pelo OpenAI treinado em cerca de 400 milhões de imagens /legendas
- Ele é treinado utilizando pares de imagens / legendas
- Clip pode ser instruído em linguagem natural para classificar uma imagem

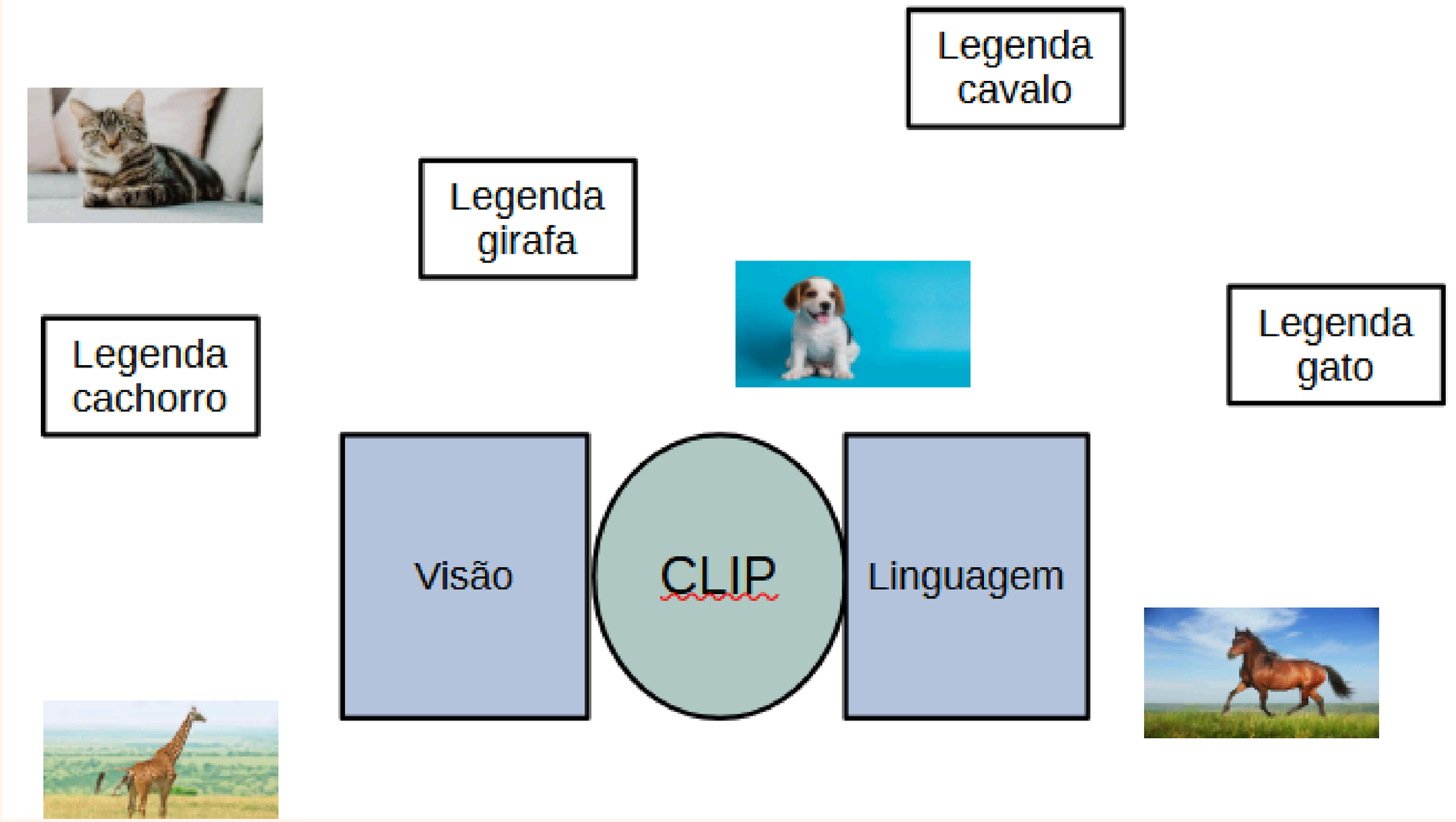


Realização: Sebrae for Startups

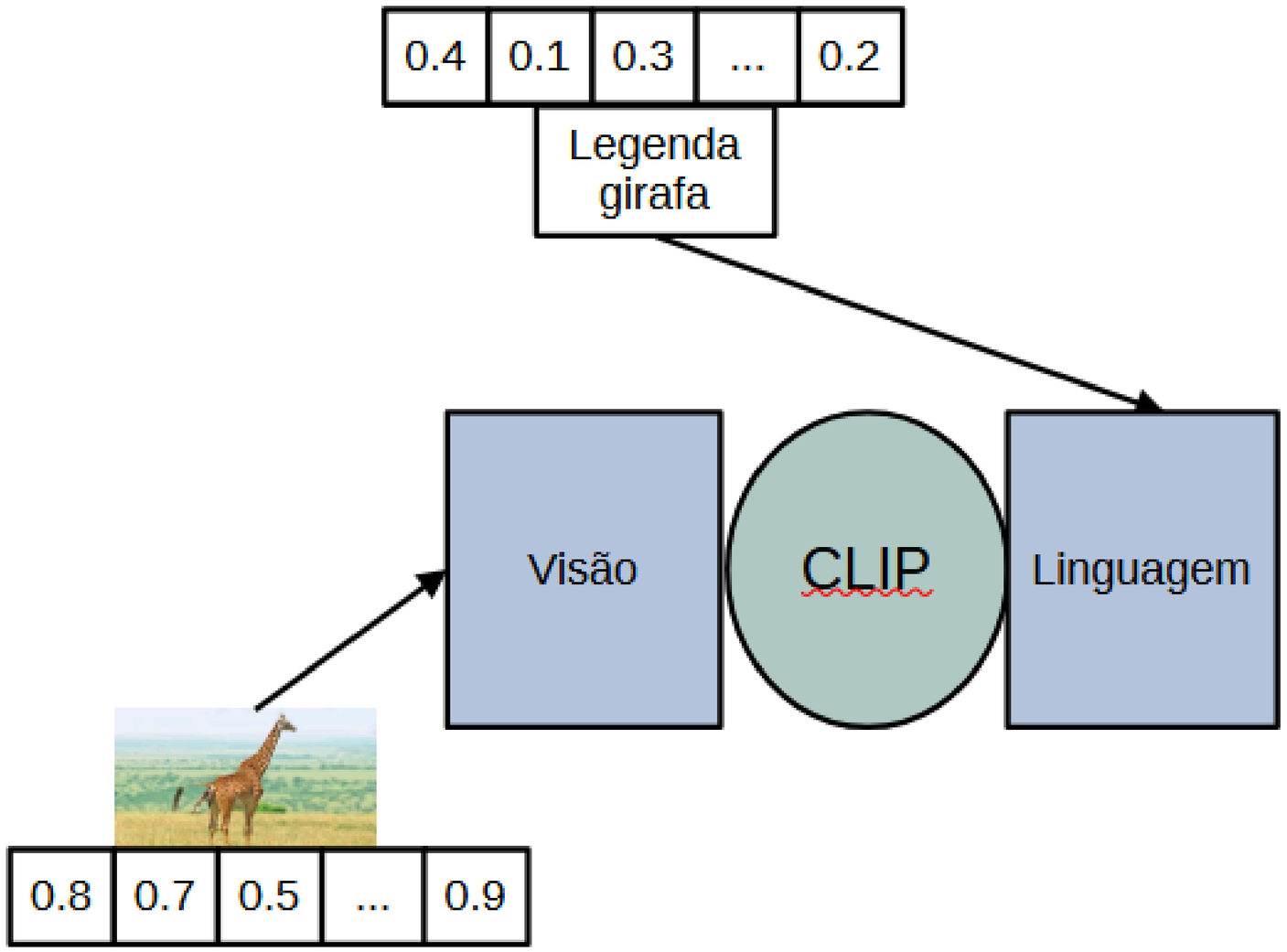
Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

Como o CLIP funciona

- Os embeddings das imagens e legendas ficam "perdidos" pelo espaço



Como o CLIP funciona

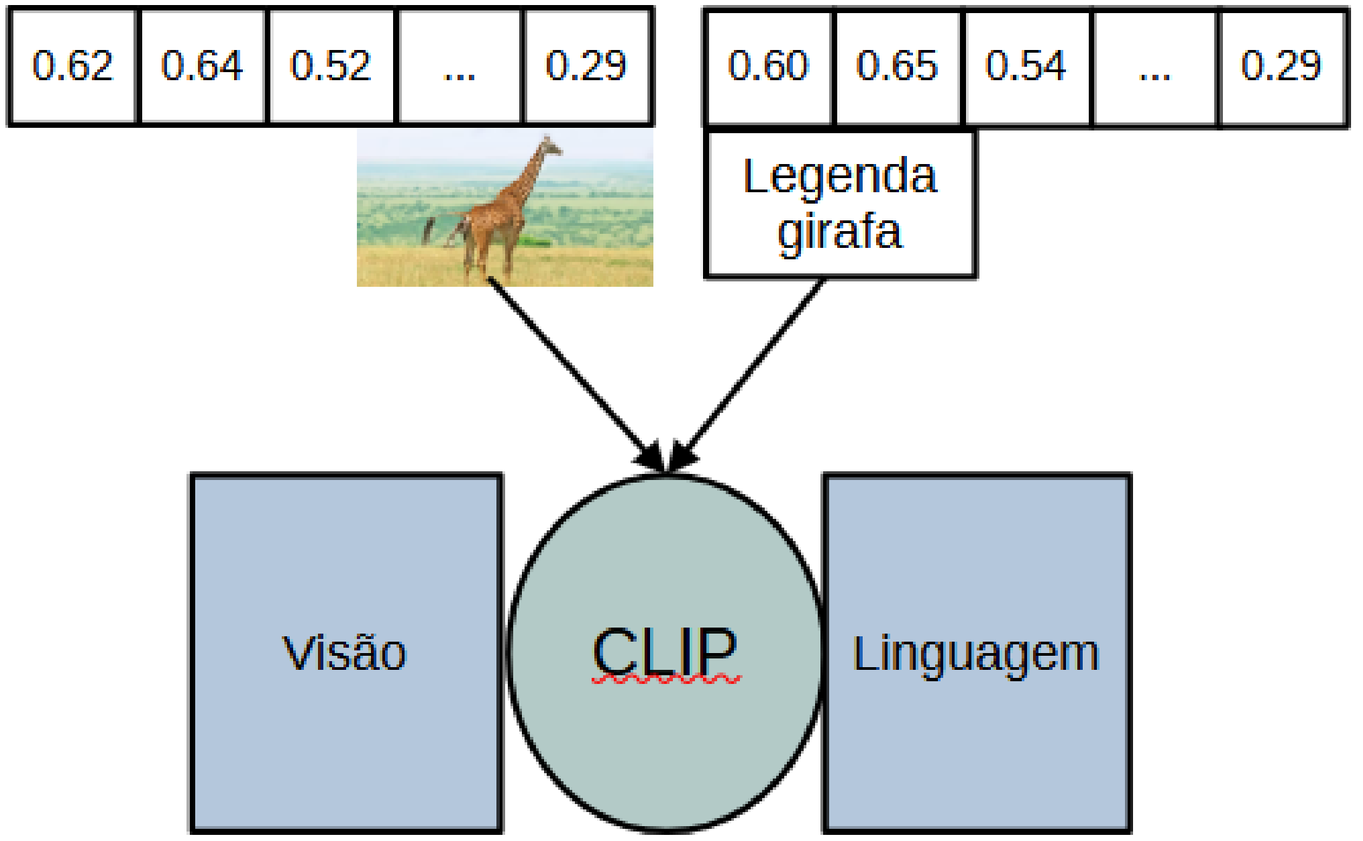


Realização:

Execução:

Como o CLIP funciona

- O trabalho do CLIP é aproximar os embeddings das imagens com os embeddings das legendas.



Realização:

Execução:

Como o CLIP funciona

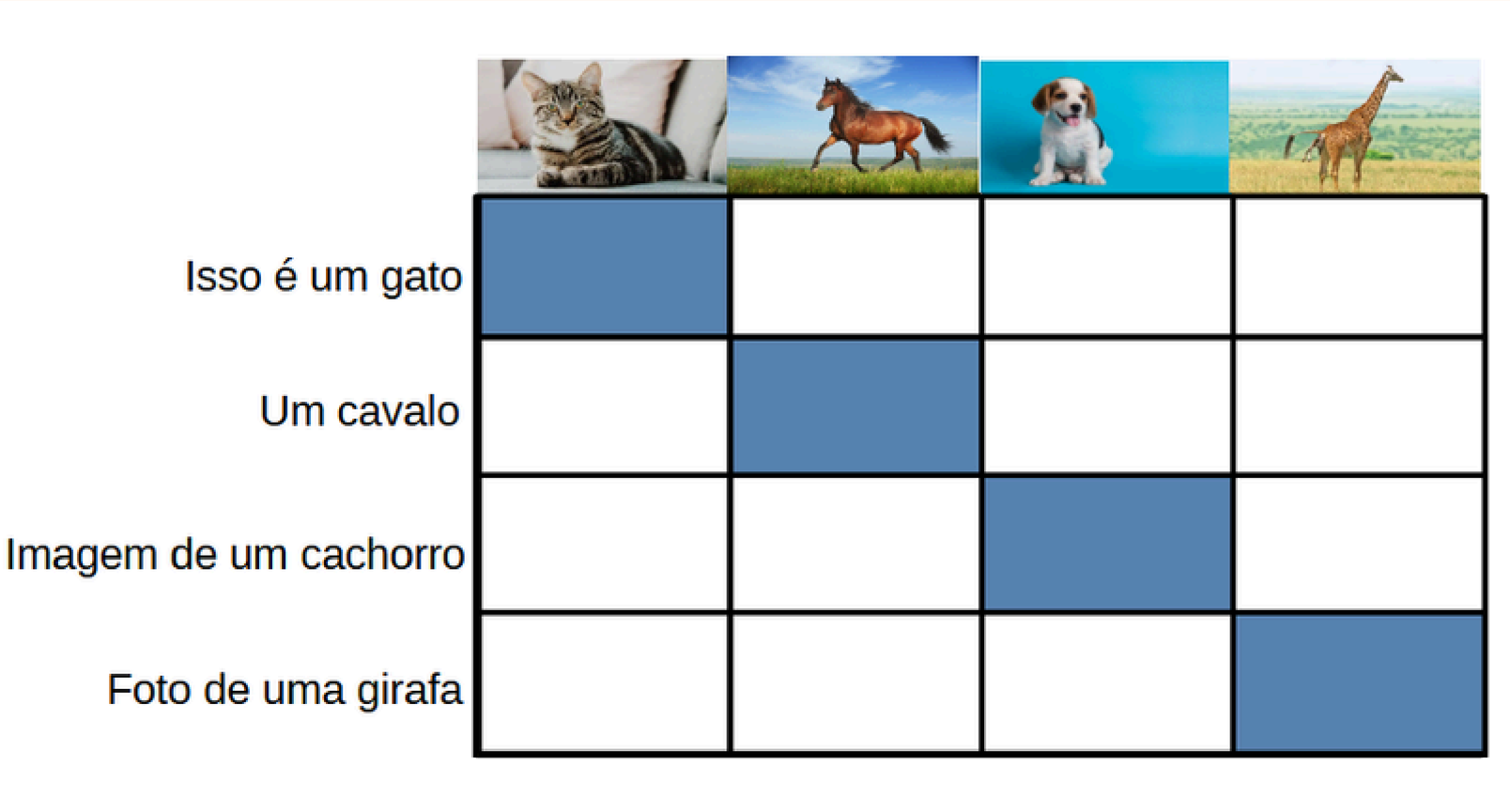
- Outra forma de ver isso

Realização:

Execução:

Como o CLIP funciona

- Outra forma de ver isso
- O CLIP **diminui** as distâncias em azul



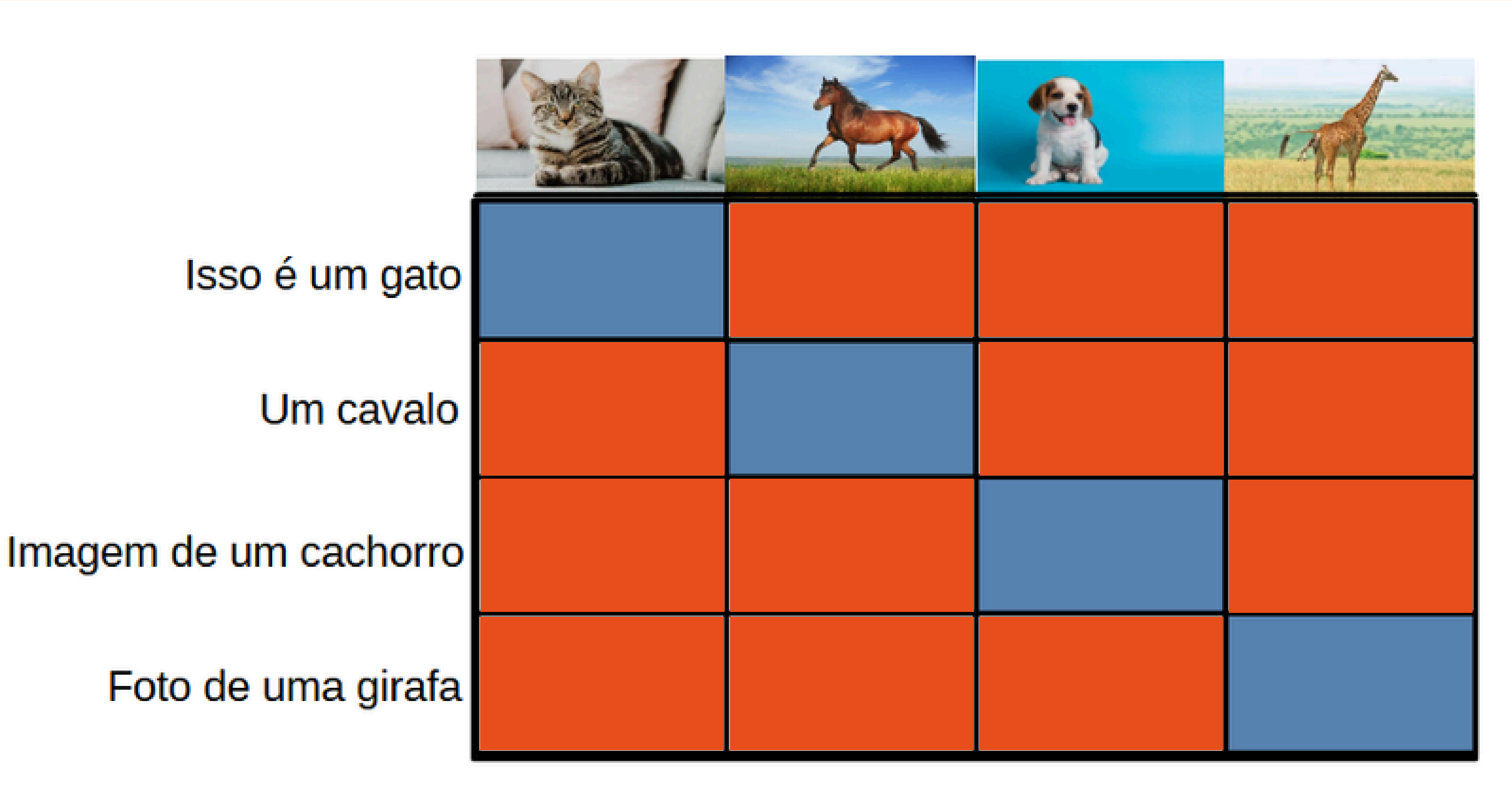
Realização: Sebrae for Startups

Execução: IPT OPEN, ipt Instituto de Pesquisas Tecnológicas

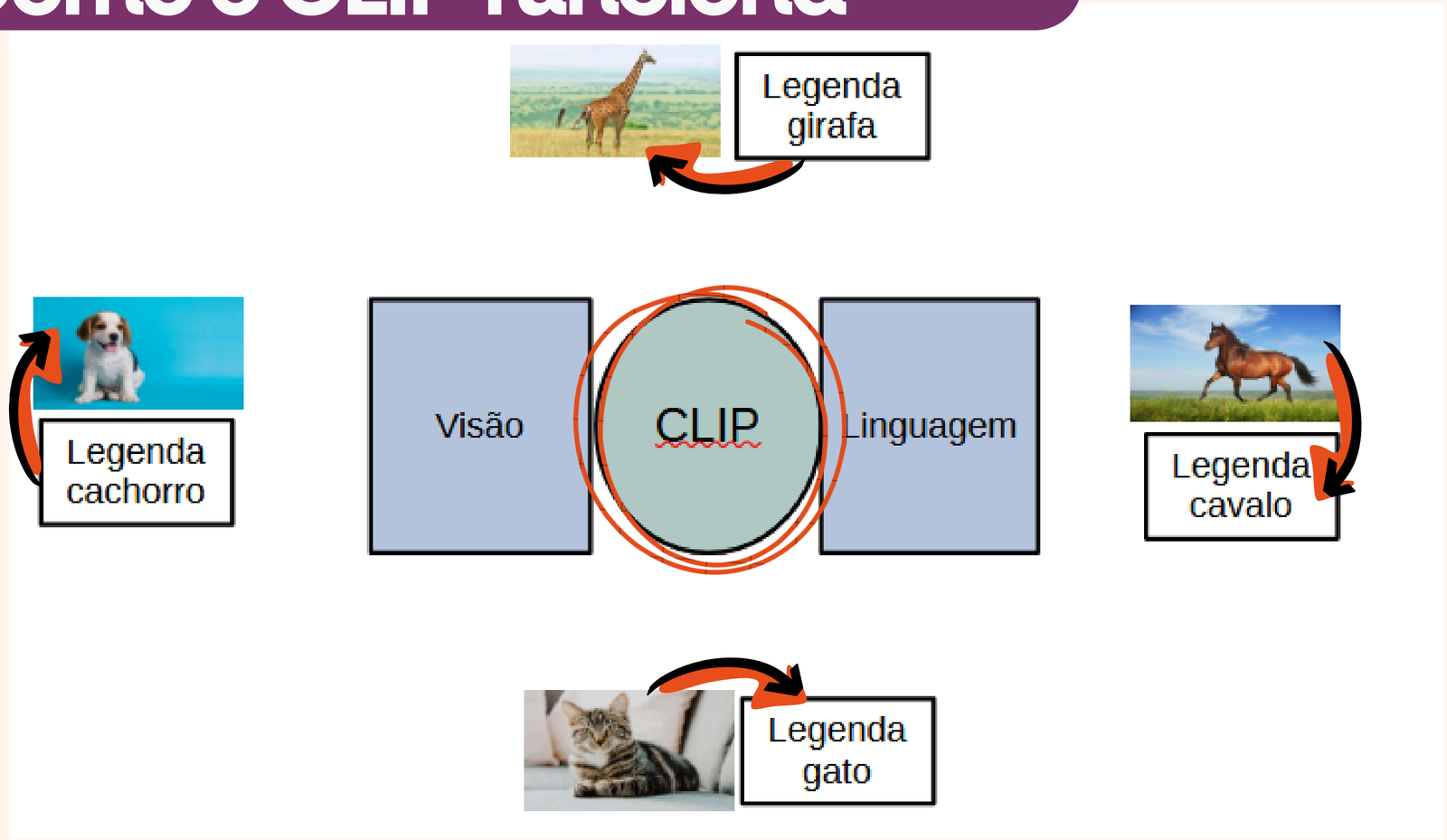
start IA

Como o CLIP funciona

- Outra forma de ver isso
- O CLIP **aumenta** as distâncias em laranja



Como o CLIP funciona



start IA

Como usar

- O CLIP deixa as imagens e as legendas correspondentes próximas.
- Mas e agora?
- Como isso nos ajuda?

Realização: Sebrae for Startups

Execução: IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Como usar

Encontrar a melhor legenda

- A legenda com os embeddings mais **próximos** é escolhida como a legenda da imagem.

Realização: Sebrae for Startups

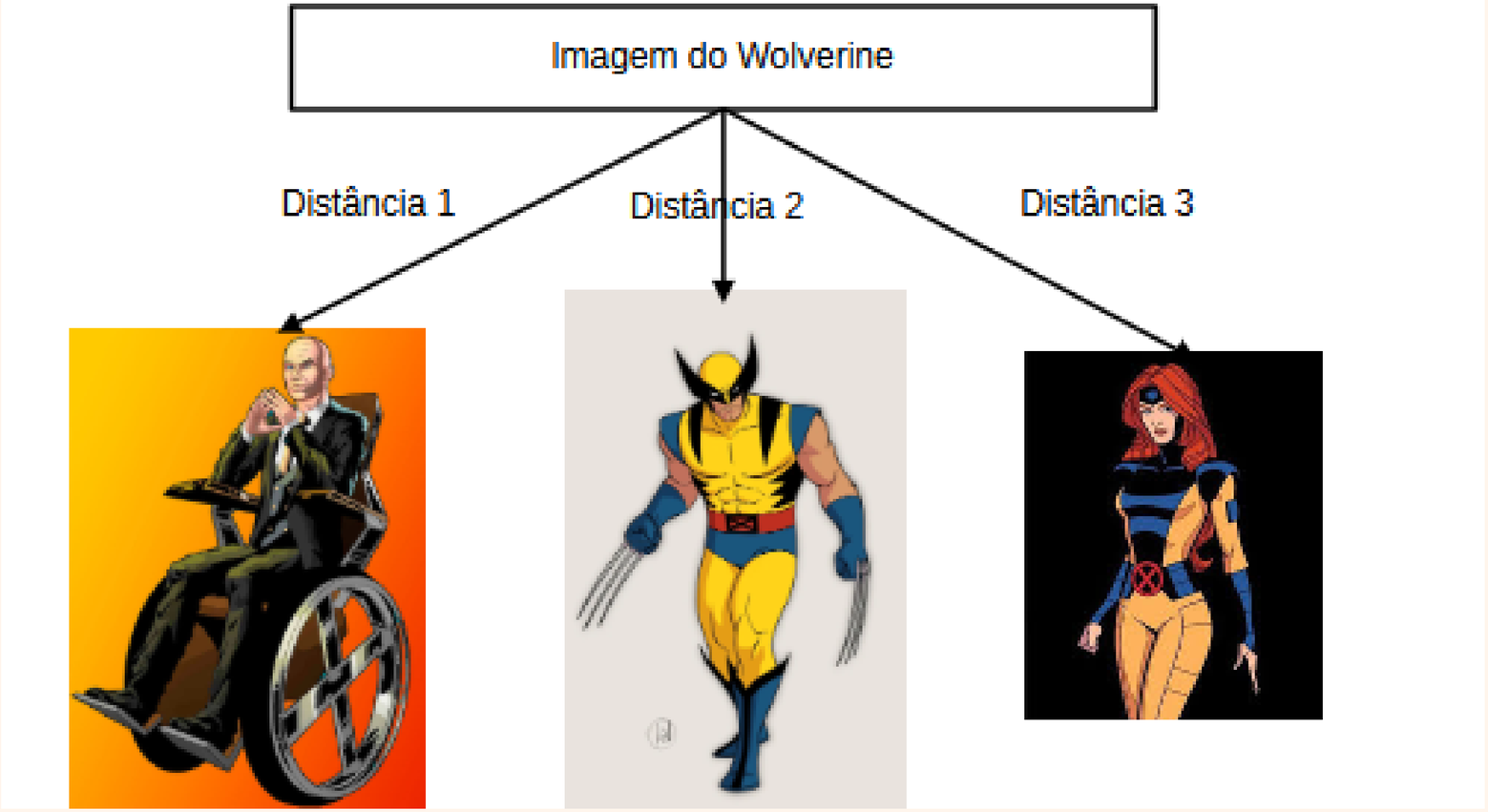
Execução: IPT OPEN, ipt Instituto de Pesquisas Tecnológicas

start IA

Como usar

Busca por imagem

- O processo contrário também é possível.
- O usuário busca por algo e as imagens com embeddings mais próximos são escolhidas.



Realização: Sebrae for Startups

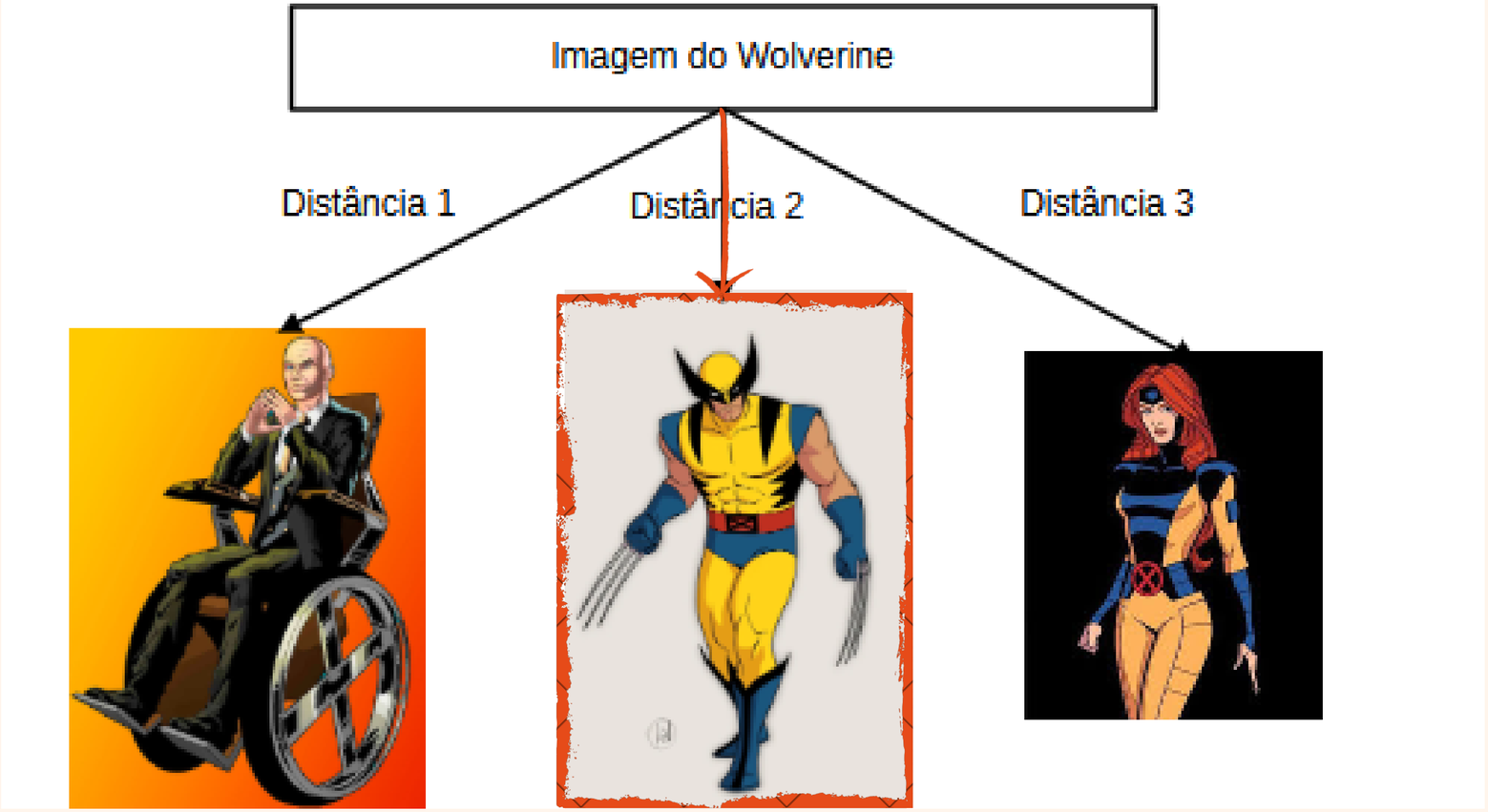
Execução: IPT OPEN, ipt Instituto de Pesquisas Tecnológicas

start IA

Como usar

Busca por imagem

- O processo contrário também é possível.
- O usuário busca por algo e as imagens com embeddings mais próximos são escolhidas.



Realização: Sebrae for Startups

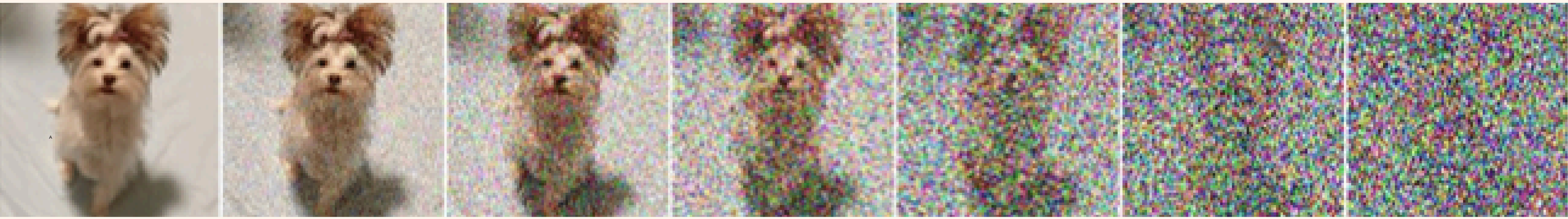
Execução: IPT OPEN, ipt Instituto de Pesquisas Tecnológicas

start IA

Mais VLMs

Modelos de difusão

- O Stable Diff usion utiliza um modelo similar ao CLIP.
- Recebe um input do usuário e gerar uma imagem correspondente.



Mais VLMs

LLaVA (Large Language and Vision Assistant)

- Conecta a visão do CLIP com o modelo de linguagem LLaMA.
- Realiza tarefas como responder perguntas

Visual input example, Extreme Ironing:

Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

User: What is unusual about this image?

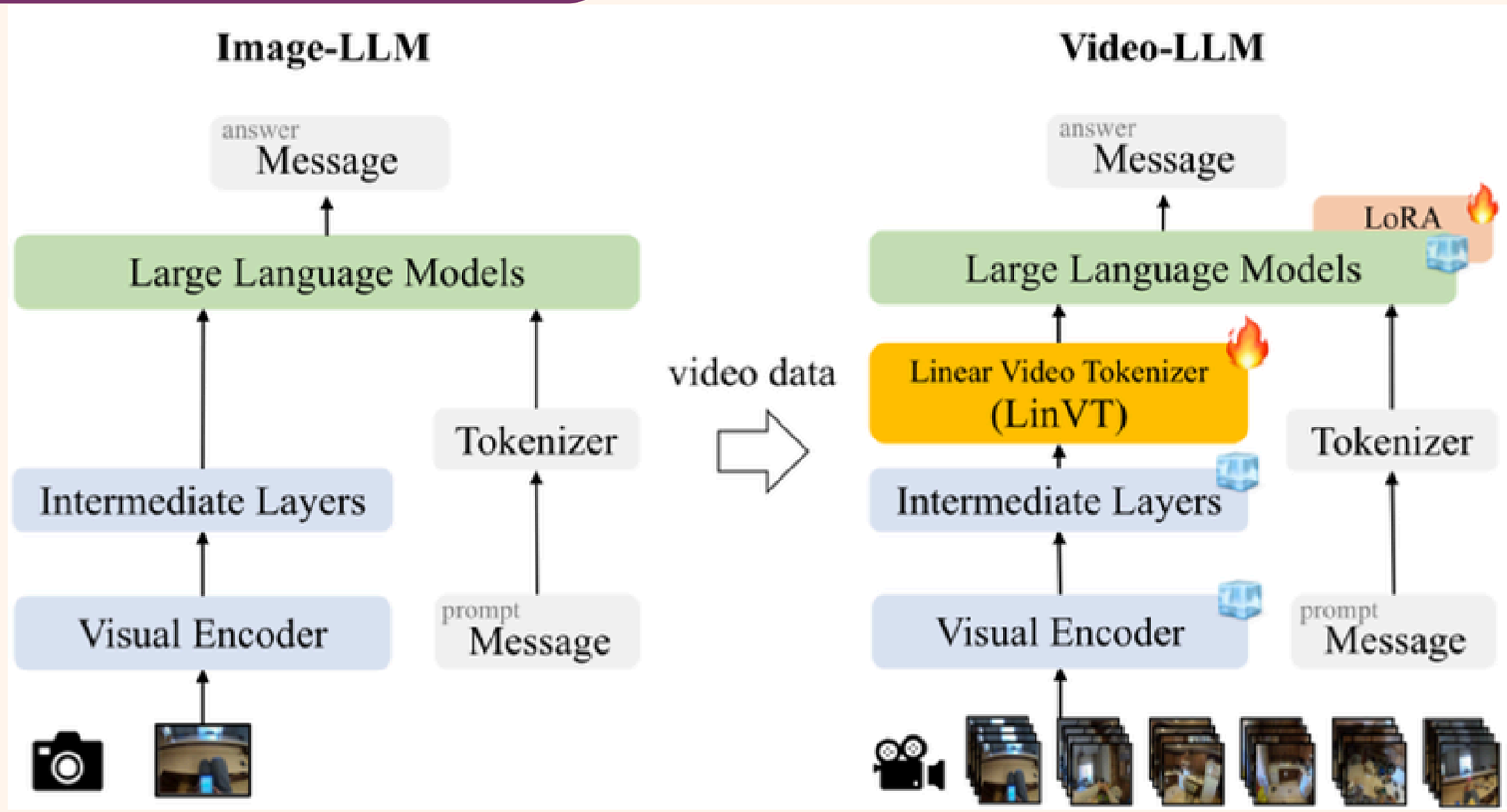
LLaVA: The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional. Additionally, it is not clear how the man is able to maintain balance and stability while ironing clothes in such an unstable environment.

start IA

Mais VLMs

VidLMs (Video Language Models)

- Modelos que podem descrever e comentar vídeos com linguagem natural.
- Capazes de lidar com dados temporais.



Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Conclusão

- O avanço dos VLMs mostra que a inteligência artifi cial está cada vez mais direcionada para tratar **múltiplos tipos de dados** simultaneamente.
- Modelos como o CLIP foram o primeiro passo, direcionando modelos a relacionar **palavras eimagens**.
- Agora, com modelos multimodais mais recentes, vemos sistemas que podem descrever, explicar e **dialogar sobre imagens e vídeos**.

Realização: Sebrae for Startups

Execução: IPT OPEN ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Obrigada!

Dúvidas?

start

Grupo de Whatsapp

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS

start IA

Apresentem-se

start IA

Agradecemos a sua participação

Um excelente ciclo!
Parabéns

Realização:

Sebrae for Startups

Execução:

IPT OPEN

ipt INSTITUTO DE PESQUISAS TECNOLÓGICAS