



# Diego Sulzbeck Villalobos

## Senior Data Scientist

Data Science • NLG • Machine Learning • Data Analytics

### CONTATO

- 📍 São Paulo, SP, Brasil
- ✉️ [diego.villa2@gmail.com](mailto:diego.villa2@gmail.com)
- ☎️ +55 11 99675-8399
- 🌐 [www.linkedin.com/in/diego-sulz/](http://www.linkedin.com/in/diego-sulz/)

### IDIOMAS

- Português - **Nativo**
- Espanhol - **Nativo**
- Inglês - **Fluente (C1)**
- Mandarim - **Iniciante**

### HABILIDADES

#### Data Science / ML

- Modelagem preditiva e scoring
- Detecção de anomalias e risco
- Séries temporais (ARIMA)
- Regressão logística
- Clustering e redução dimensional
- Bayes e árvores (C4.5)
- Redes neurais; Random Forest
- Tratamento de imagem
- RAG (BERT, XLM-R, E5)
- Fine-tuning (QLoRA, LLaMA)

#### Ferramentas / Stack

- Python, SQL, R
- Power BI, Tableau
- Excel
- Splunk
- Java, JavaScript
- (Eng. de Software e Segurança)

### REPOSITÓRIOS

- 🤖 [huggingface.co/DiegoVSulz](https://huggingface.co/DiegoVSulz)
- 🐙 [github.com/DiegoSVilla](https://github.com/DiegoSVilla)

### INFORMAÇÕES PESSOAIS

- 33 anos (1992)
- Dupla cidadania(ESP-BR)
- 5 anos morando no Chile
- Amante de jogos de tabuleiro
- 5 cachorros maravilhosos

### IA GERATIVA

#### Refinamentos QLoRA (LLaMA2):

Capivarinha7b; Capivara13b.

**RAGs:** BERT + embeddings (XLM-

#### Engenheiro da Computação (Poli-USP)

**9 anos** de experiência em **Dados e Segurança**

**7 anos** de experiência como **Data Scientist**

**3 anos** de especialização em **NLP, NLG e IA Gerativa**

### TECNOLOGIAS

**Python:** Scikit-learn, Torch, Pandas, TensorFlow, HuggingFace, OAI, entre várias outras

**Data Vis:** PBI, Tableau, Matplotlib, Seaborn, Plotly

**Agents and MAS:** LangGraph, LangChain, Flowise, N8N, PydanticAI

**MLOps:** wandb, MLFlow, HuggingFace

**Cloud:** AWS, Azure, Google Cloud, entre outras, principalmente a nível de usuário

### OBJETIVO

**Seguir atuando como referência técnica** em times que transformam **soluções de Ciência de Dados Gerativa** em solução produtivas de ponta, com rigor técnico e aderência a padrões e tempos de mercado **com altíssima competitividade**

### HISTÓRICO PROFISSIONAL

#### Future Security AI - Senior Data Scientist

2025 - Atual

- Modelos baseados em Transformers para solução de desafios de voz com alta complexidade lógica.
- Flowise e N8N para construção de sistemas multi-agentes (MAS) e pipelines de automação de trabalho utilizando agentes.
- Pipelines de modelos e prompts para manutenção de soluções MAS produtivas utilizando MLFlow.
- Avaliação seguindo práticas de DS de modelos, prompts e agentes produtivos, com pipelines automatizadas.
- 3 meses na Austrália para treinamento de recursos de DS em clientes estratégicos.

#### Houer - Senior Data Analyst

2023 - 2025

- Uso de NLP e NLG (Llama, OpenAI, Bert) para a automatização de relatórios e consultas de dados com recorrência temporal.
- RAG de dados para visualização em tempo real de projetos com milhões de colunas.
- Relatórios, POCs e dashboards (Data Analytics + BI).
- Inteligência de negócio e acompanhamento de indicadores para PPPs.
- Manutenção de ETLs utilizando pipelines AWS

Roberta, E5).

### Projetos:

- **MonitorIA:** LLaMA fine-tuned; VDB + RAG com grandes bases de conteúdo.
- **Mordomo Fantasma:** Whisper (voz→texto) + Llama (texto→texto) + Bark (texto→voz) + RAG.
- **Wiki de Projeto:** documentação e Q&A técnico via Vector DB.

## HACKATHON

### Vitória Hackathon - Ambev 2015

- Desenvolvimento (Ionic / Eng. de Software); arquitetura e engenharia.
- Projeto acelerado; pitch para investidores; visita ao Vale do Silício.

### LexisNexis Risk - Senior Data Scientist

2022 – 2023

- Uso de técnicas de Ciência de Dados para identificação e modelagem de soluções de ML visando a identificação de agentes mal intencionados
- Estudo do cenário Brasileiro de fraude bancária com o intuito de manter a solução atualizada para o mercado nacional e identificar MOs emergentes.
- Interface com clientes para apresentar resultados de modelos de dados, próximos passos e melhores práticas.

### BioCatch - Threat Analyst for Latin America

2019 – 2021

- Data Analytics + Data Science; SQL e Python.
- Relatórios, POCs e customizações de ML para segurança bancária.
- Experiência internacional: empresa israelense atendendo mercado LatAm.

### Bradesco - Security and Innovation Engineer

2018 – 2019

- Modelos preditivos para inovação.
- Modelos de supervisionados e não supervisionados identificação de malware: lustering, redução dimensional, tratamento de imagem, Bayes, RN, RF, C4.5

### Itaú - Security Analyst

2015 – 2017

(Java Development)

- Segurança em canais digitais; supervisão de projetos.
- Métricas de fraude e BI; ferramentas terceiras; gestão; programa trainee.

## FORMAÇÃO

### Mestrado SC- Poli-USP

2025 – Em andamento

Orientador: Prof. Dr. Bruno C. de Souza Sanches

Tema: Arquitetura ASIC para modelos baseados em Transformers visando aplicação robótica

### Engenharia Elétrica (Computação) - Poli-USP

2012 – 2022

(Bacharelado)

TCC: Protocolo de redes IoT focado em segurança, descentralização, resiliência a falhas e ataques.