



GT Malware DataLab

<https://malwaredatalab.github.io>



Demo Day



29 de novembro de 2024

Equipe



Diego Kreutz

Experiência em Projetos de
Malware Android e formação
de recursos humanos



Rodrigo Mansilha

Experiência em Projetos de
IA generativa e formação
de recursos humanos



Hendrio Bragança

Experiência em IA, Ciência de
Dados e Soluções para
malware Android



Kayuã Paim

Experiência em IA Generativa
e Desenvolvimento de
Software



Luiz Laviola

Experiência em Sistemas,
Engenharia e
Desenvolvimento de Software



Angelo Diniz

Experiência em Sistemas,
Infraestrutura e
Desenvolvimento de Software



Colaboradores



Anna Luiza Gomes

Profissional de Ciência de Dados
com Experiência de Mercado
Mestranda em Eng. de Software



Leonardo Sonco

Profissional de Front End
Graduando em Ciência da
Computação



Lucas Ferreira

Profissional de Cibersegurança
com Experiência de Mercado
Mestrando em Eng. de Software



Rafael Nogueira

Voluntário de Iniciação Tecnológica
Graduando em Ciência da
Computação

Malware Android: o problema persiste



Android malware is wiping out bank accounts in Finland

Updated on: May 06, 2024 8:55 AM



Paulina Okunytė, Journalist



**Usuários
perdem várias
dezenas de
milhares de
EUROS**

Como resolver o problema?

Google Unveils Advanced AI Malware Detection for Android



By Sophia Taylor



Inteligência Artificial (IA)
Avançada é necessária
para combater malwares



Mas por que projetos de IA falham?

14 Reasons Why 70 to 80% of AI projects fail: Decoding the High Failure Rate of AI Projects



Kenneth Nicholaus · Follow

5 min read · Jul 12, 2024



Dados insuficientes
ou de baixa
qualidade e falta de
cultura orientada a
dados. IA é centrada
nos dados!
Os datasets
existentes são
defasados e
limitados!

Como obter datasets de qualidade?

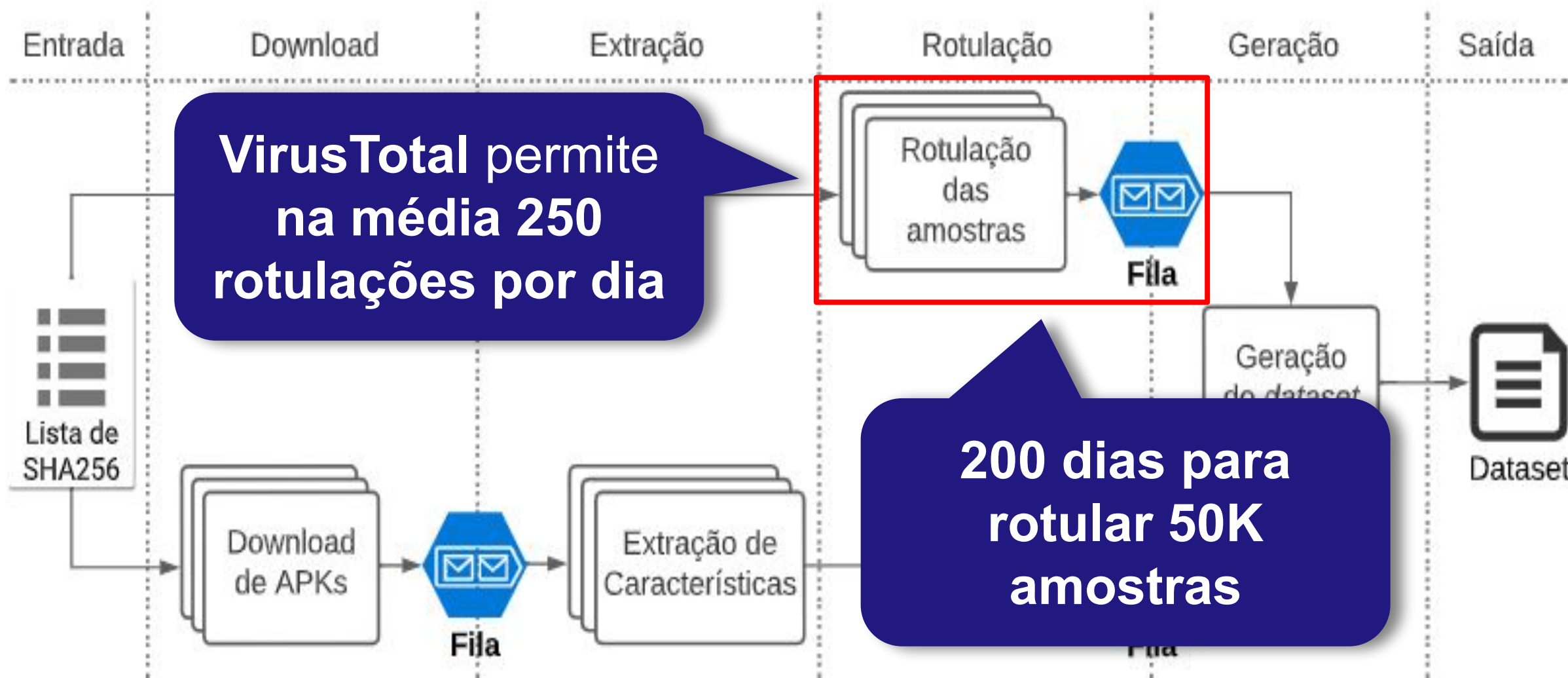
ADBuilder



AMGenerator e AMExplorer



ADBuiIer



Como resolver o problema?

Forbes

FORBES > INNOVATION > ENTERPRISE TECH

20 Generative AI Tools For Creating Synthetic Data

Bernard Marr Contributor

Follow



0

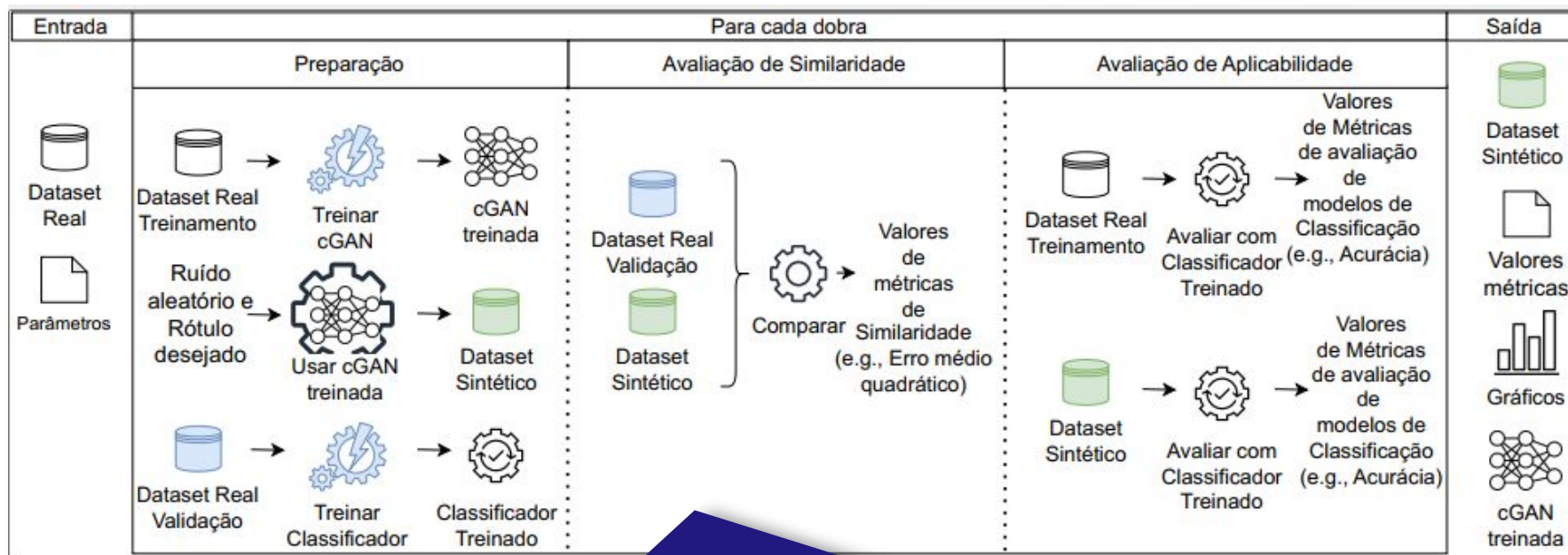


Aug 29, 2024, 02:51am EDT



Uma resposta: Ferramentas
de IA Generativa para
Geração de Dados Sintéticos

DroidAugmentor (2023)



1. **preparação:** treinamento da cGAN
2. **avaliação (similaridade):** métricas
3. **avaliação (aplicabilidade):** métricas



DroidAugmentor: uma ferramenta de treinamento e avaliação de cGANs para a geração de dados sintéticos

Karina Casola, Kayuã Oleques Paim, Diego Kreutz, Rodrigo Brandão Mansilha

XXIII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais

https://sol.sbc.org.br/index.php/sbseg_estendido/article/view/27273

AutoDroid (2023)



- Código Open Source (GitHub)
- Desenvolvido em Typescript para Node.js
- Utilização através de API REST

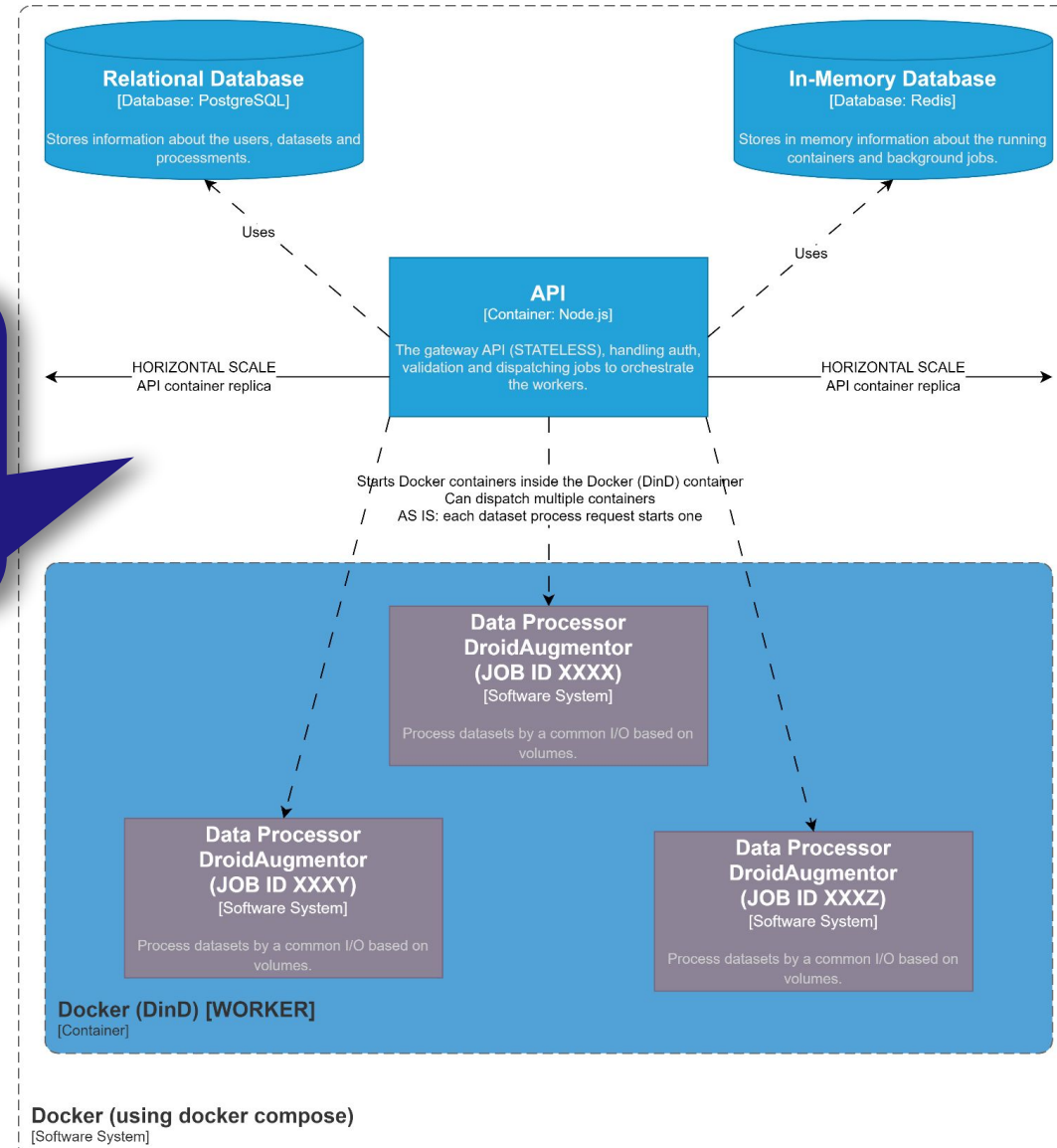


AutoDroid: disponibilizando a ferramenta DroidAugmentor como serviço

Luiz Felipe Laviola, Kayuã Oleques Paim, Diego Kreutz, Rodrigo Brandão Mansilha

XX Escola Regional de Redes de Computadores (ERRC 2023)

<https://sol.sbc.org.br/index.php/errc/article/view/26020>



Hackers do Bem (2023)



**GT - Malware DataLab: Serviço de
Execução e Avaliação de Redes Neurais
Artificiais para Geração de Dados
Sintéticos de Malware**

Laboratório de Experimentação (produto 1)

Oferecer um serviço para reduzir a curva de aprendizado e facilitar a investigação de técnicas avançadas de geração de dados sintéticos para expandir *datasets* úteis na classificação de aplicativos Android (benigno ou maligno).

- O1. Ambiente de Execução (extensão inovadora)
- O2. Ambiente de Avaliação (pesquisa)
- O3. Material de Aprendizado e Treinamento (ensino)

Laboratório de Ensino-Aprendizagem (produto 2)

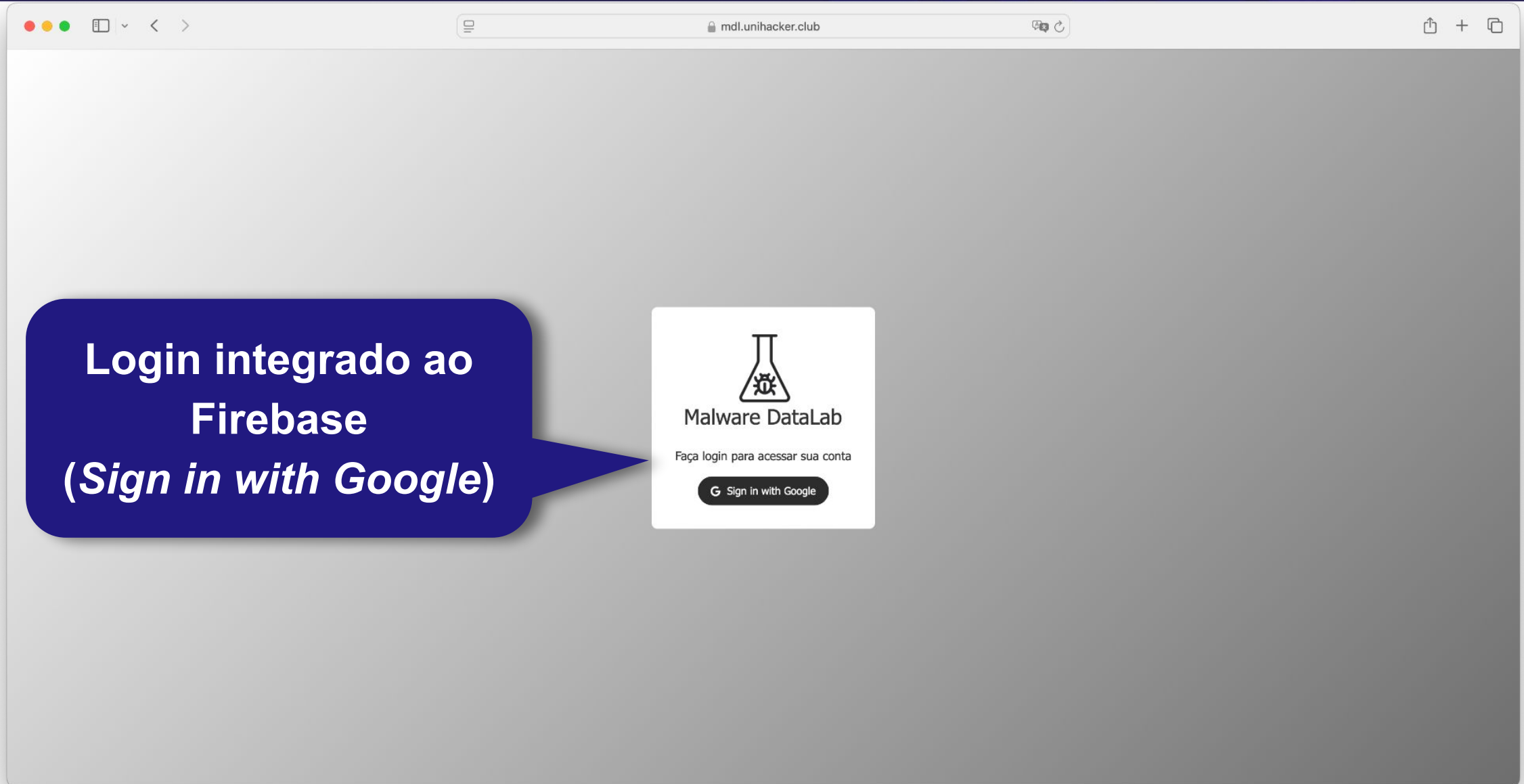
Capacitar os Hackers do Bem através de uma plataforma de Ensino-Aprendizagem para formação de recursos humanos qualificados em IA Generativa para gerar dados sintéticos.

- Módulos com material teórico e prático
- Avaliação diagnóstica e avaliações somativas
- Introdução a laboratórios adicionais de aprendizagem

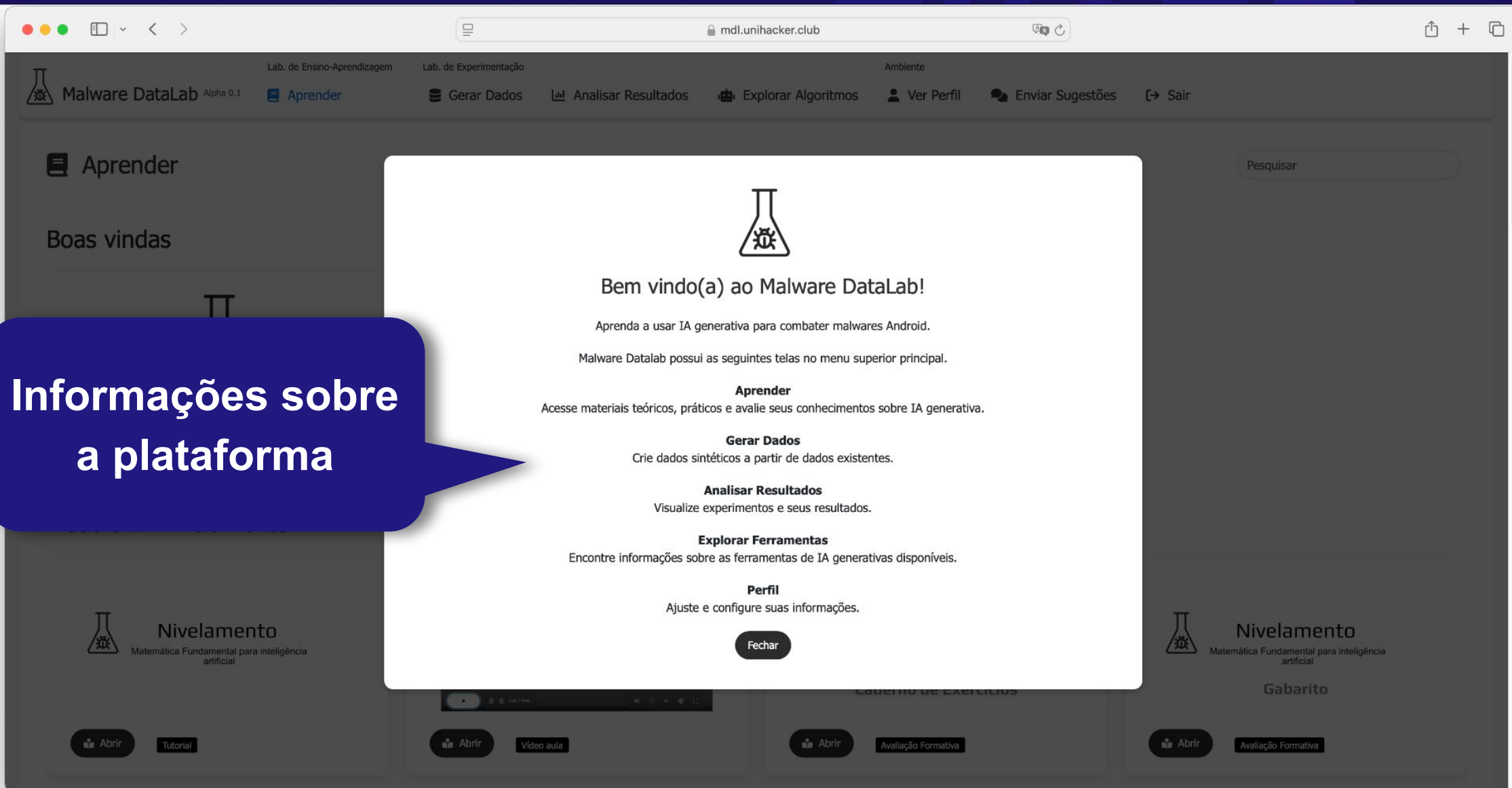
1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. MalSynGen (evolução da DroidAugmentor)
4. Cloud AutoDroid (evolução da AutoDroid)
5. GUI4MalSynGen (versão minimalista)

1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. MalSynGen (evolução da DroidAugmentor)
4. Cloud AutoDroid (evolução da AutoDroid)
5. GUI4MalSynGen (versão minimalista)

Malware DataLab: Login



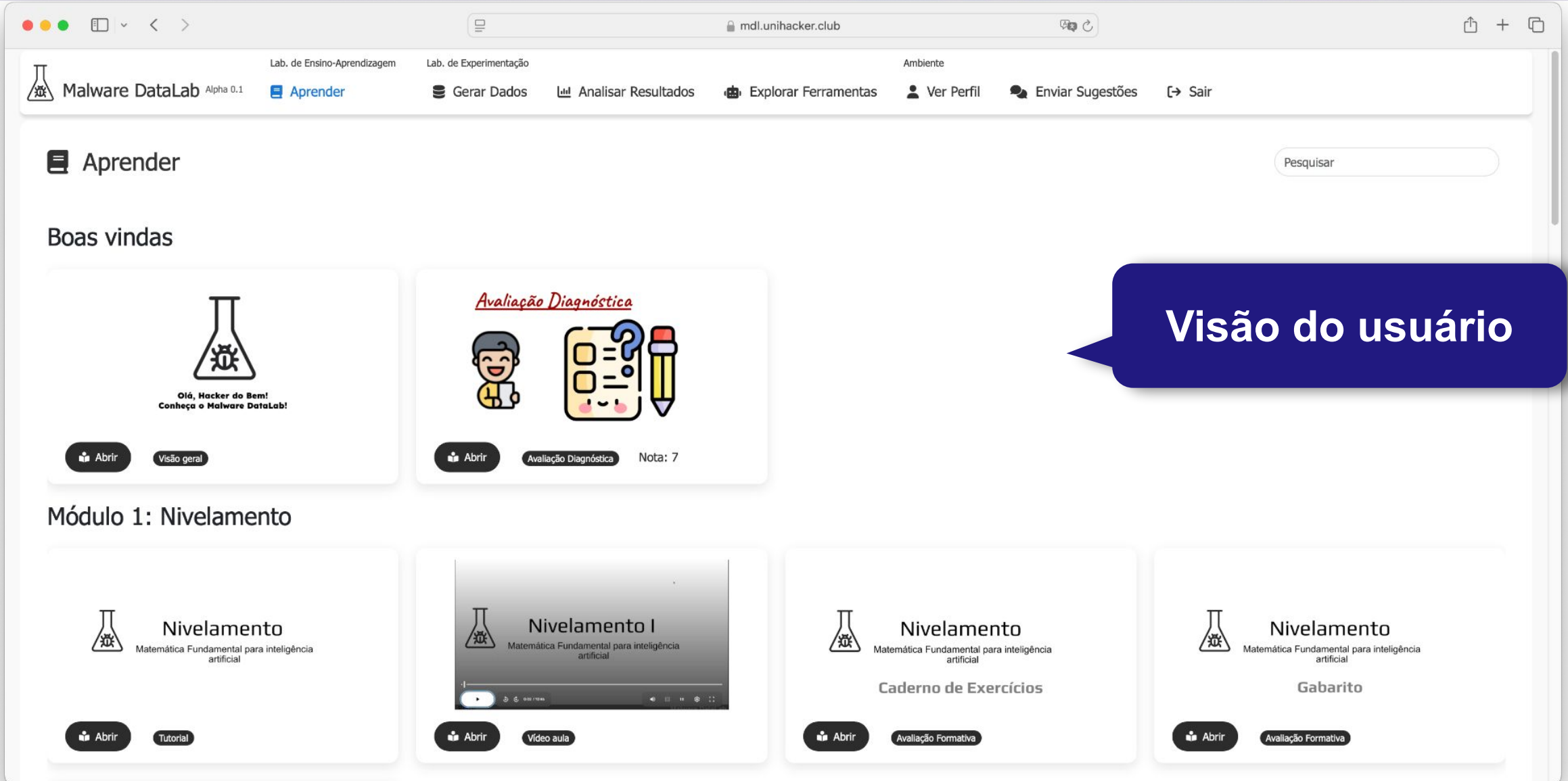
Malware DataLab: Boas vindas



The screenshot shows a web browser window with the URL `mdl.unihacker.club`. The application header includes a navigation bar with links: **Malware DataLab** Alpha 0.1, **Aprender**, **Gerar Dados**, **Analisar Resultados**, **Explorar Algoritmos**, **Ver Perfil**, **Enviar Sugestões**, and **Sair**. Below the header, a sidebar on the left contains **Aprender** and **Boas vindas**. A central modal window displays the welcome message: **Bem vindo(a) ao Malware DataLab!** and provides instructions on using generative AI to combat Android malware. The modal lists five main features: **Aprender** (access to theoretical and practical materials), **Gerar Dados** (create synthetic data), **Analisar Resultados** (visualize experiments), **Explorar Ferramentas** (find generative AI tools), and **Perfil** (adjust user information). A **Fechar** button is at the bottom of the modal. The background interface also shows sections for **Nivelamento** (Fundamental Mathematics for Artificial Intelligence) and **Gabarito** (Answer Key).

Informações sobre a plataforma

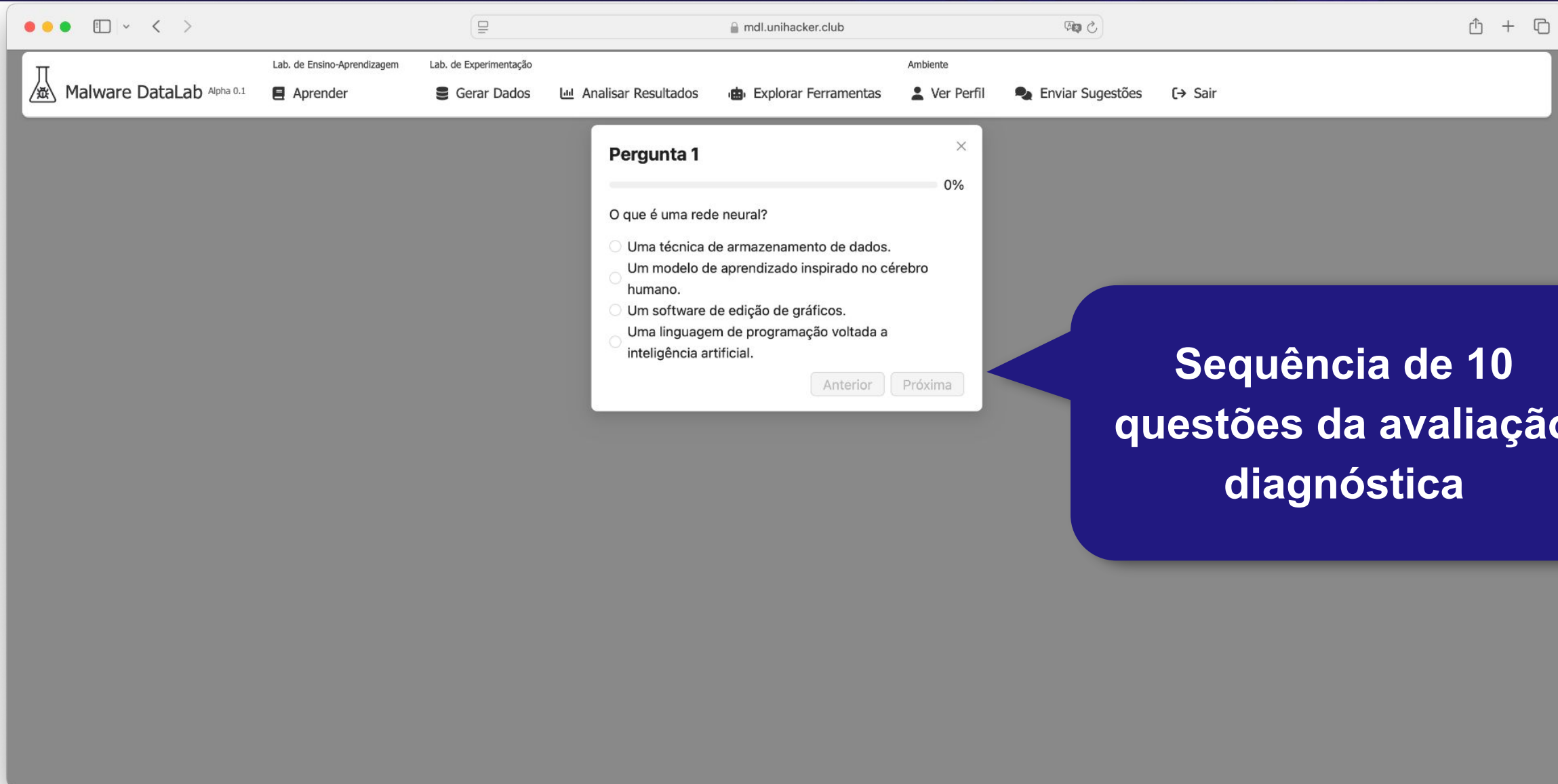
Lab. de Ensino-Aprendizagem



The screenshot displays the Malware DataLab web application interface. At the top, a navigation bar includes the site logo, the name 'Malware DataLab Alpha 0.1', and several menu items: 'Aprender', 'Gerar Dados', 'Analisar Resultados', 'Explorar Ferramentas', 'Ver Perfil', 'Enviar Sugestões', and 'Sair'. Below the navigation bar, the main content area is titled 'Aprender' and features a search bar. The page greets the user with 'Boas vindas'. There are two main interactive cards: one for 'Avaliação Diagnóstica' (Diagnostic Assessment) showing a user profile and a score of 7, and another for 'Módulo 1: Nivelamento' (Module 1: Leveling). The 'Módulo 1: Nivelamento' section contains four sub-cards: 'Nivelamento' (with a tutorial button), 'Nivelamento I' (with a video lesson button), 'Nivelamento' (with a formatative assessment button), and 'Nivelamento' (with a formatative assessment button). Each card includes an 'Abrir' (Open) button.

Visão do usuário

Avaliação diagnóstica



The screenshot shows a web browser window with the URL `mdl.unihacker.club`. The application header includes a navigation bar with the following items: **Malware DataLab** Alpha 0.1, **Lab. de Ensino-Aprendizagem**, **Lab. de Experimentação**, **Ambiente**, **Aprender**, **Gerar Dados**, **Analisar Resultados**, **Explorar Ferramentas**, **Ver Perfil**, **Enviar Sugestões**, and **Sair**. The main content area displays a modal window titled **Pergunta 1** with a progress bar at 0%. The question is: "O que é uma rede neural?". There are four radio button options: "Uma técnica de armazenamento de dados.", "Um modelo de aprendizado inspirado no cérebro humano.", "Um software de edição de gráficos.", and "Uma linguagem de programação voltada a inteligência artificial.". At the bottom of the modal are buttons for **Anterior** and **Próxima**.

Pergunta 1

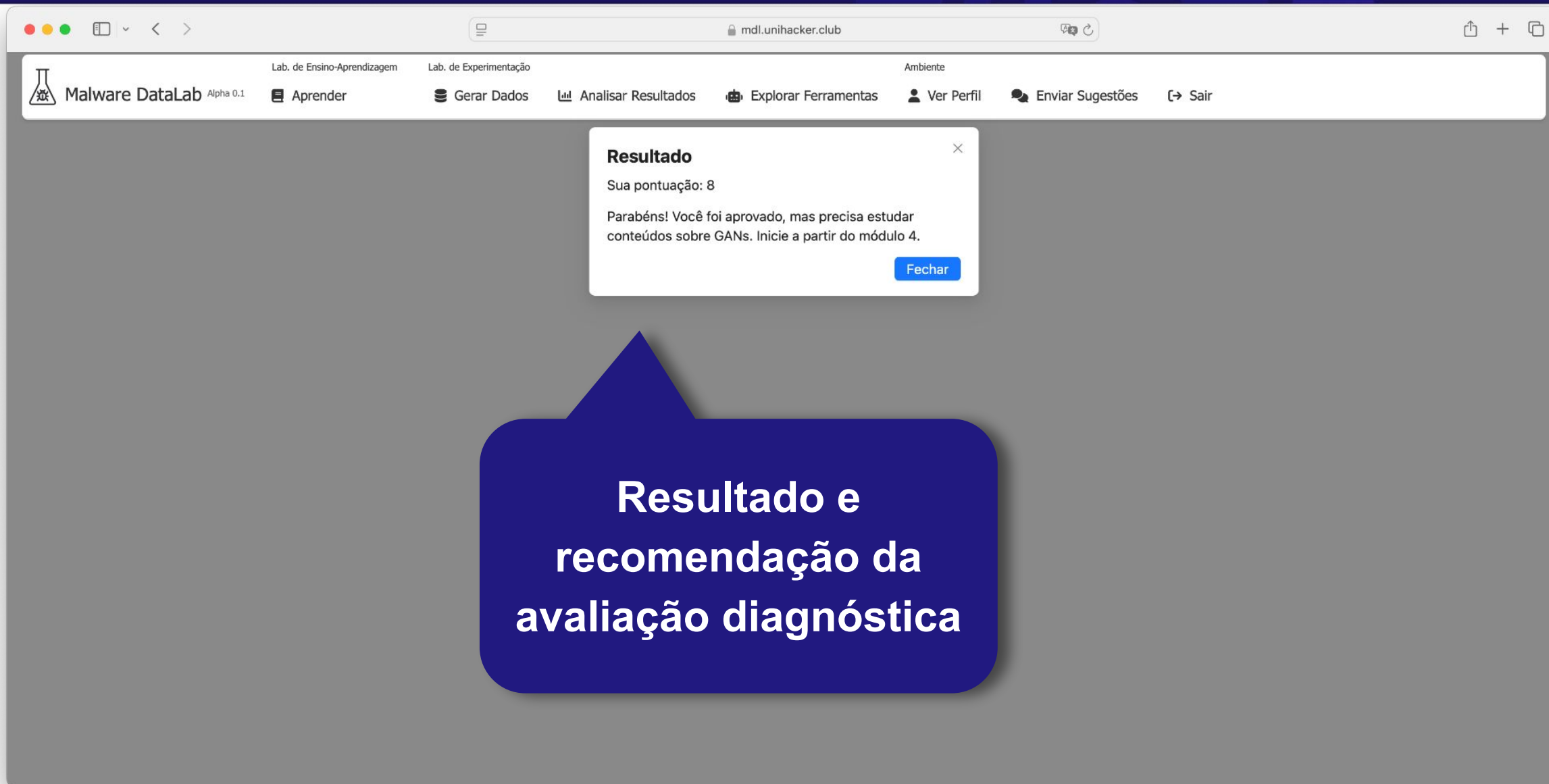
0%

O que é uma rede neural?

- ☐ Uma técnica de armazenamento de dados.
- ☐ Um modelo de aprendizado inspirado no cérebro humano.
- ☐ Um software de edição de gráficos.
- ☐ Uma linguagem de programação voltada a inteligência artificial.

Sequência de 10 questões da avaliação diagnóstica

Diagnóstico



The screenshot shows a web browser window with the URL `mdl.unihacker.club`. The application header includes the 'Malware DataLab Alpha 0.1' logo and navigation links: 'Lab. de Ensino-Aprendizagem', 'Lab. de Experimentação', and 'Ambiente'. Below these are buttons for 'Aprender', 'Gerar Dados', 'Analisar Resultados', 'Explorar Ferramentas', 'Ver Perfil', 'Enviar Sugestões', and 'Sair'. A modal window titled 'Resultado' is displayed in the center, showing a score of 8 and a recommendation to study GANs. A large blue callout box is overlaid on the bottom half of the screen.

Resultado

Sua pontuação: 8

Parabéns! Você foi aprovado, mas precisa estudar conteúdos sobre GANs. Inicie a partir do módulo 4.

[Fechar](#)

**Resultado e
recomendação da
avaliação diagnóstica**

Módulos

Módulo 1: Nivelamento

1. Slides
2. Vídeo
3. Caderno de exercícios
4. Gabarito
5. Avaliação somativa



Malware DataLab Alpha 0.1

Lab. de Ensino-Aprendizagem

Lab. de Experimentação

Aprender

Gerar Dados

Aprender

Módulo 1: Nivelamento

Nivelamento

Matemática Fundamental para inteligência artificial

Abrir

Tutorial

Nivelamento I

Matemática Fundamental para inteligência artificial

Abrir

Vídeo aula

Nivelamento

Matemática Fundamental para inteligência artificial

Caderno de Exercícios

Abrir

Avaliação Formativa

Nivelamento

Matemática Fundamental para inteligência artificial

Gabarito

Abrir

Avaliação Formativa

Avaliação Somativa

Abrir

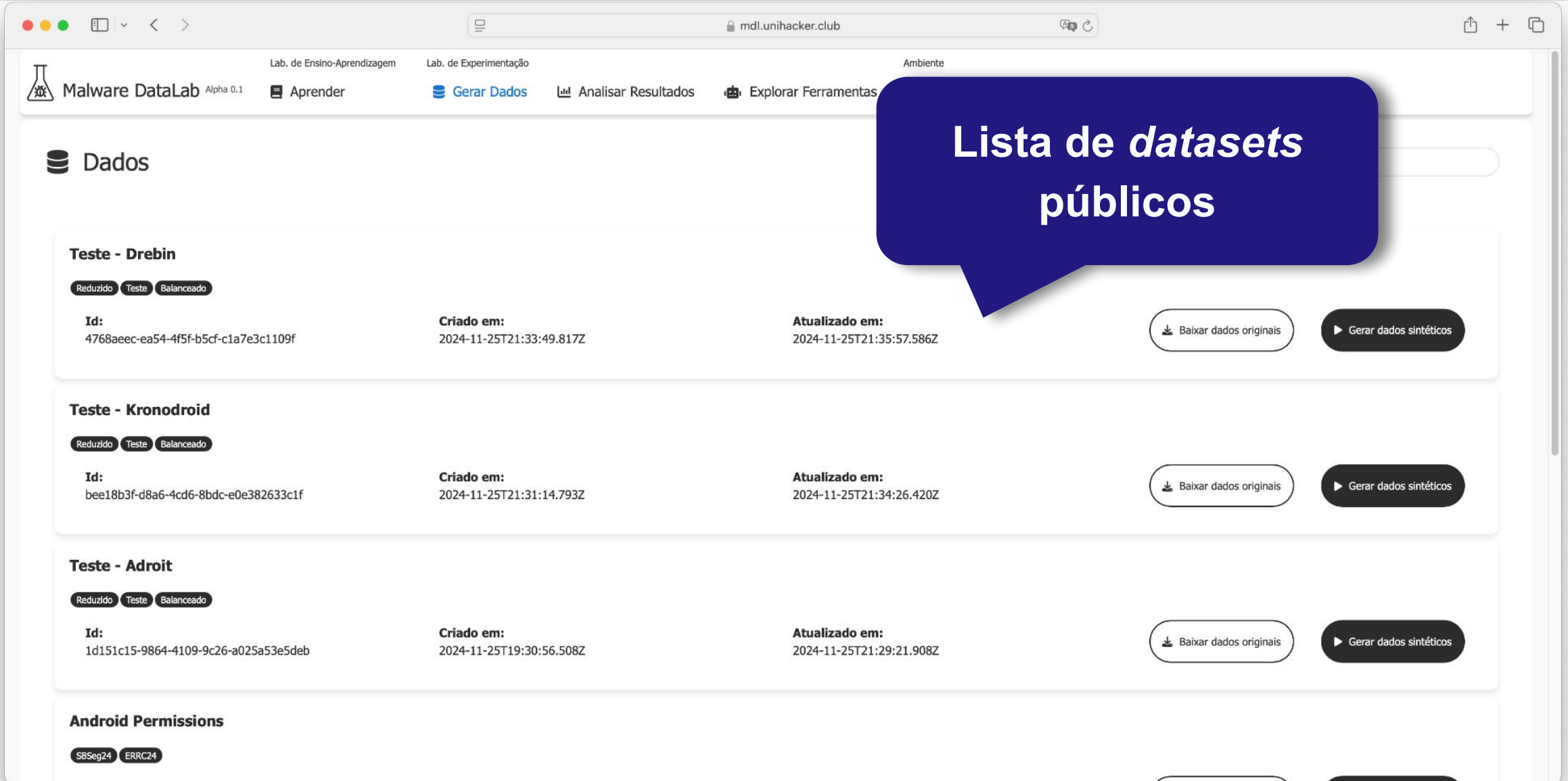
Avaliação Somativa

Demo Day Malware DataLab



1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. MalSynGen (evolução da DroidAugmentor)
4. Cloud AutoDroid (evolução da AutoDroid)
5. GUI4MalSynGen (versão minimalista)

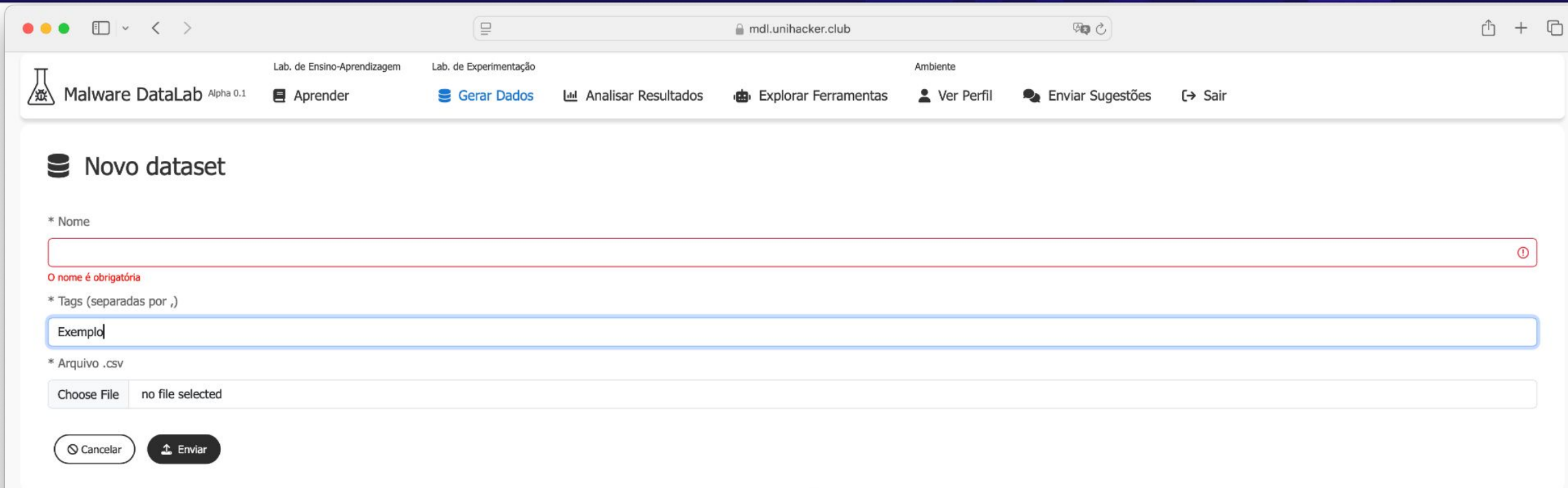
Lab. de Experimentação: Dados



The screenshot shows the Malware DataLab web application interface. The browser address bar displays 'mdl.unihacker.club'. The application header includes a navigation menu with 'Lab. de Ensino-Aprendizagem', 'Lab. de Experimentação', and 'Ambiente'. Below the header, there are tabs for 'Aprender', 'Gerar Dados', 'Analisar Resultados', and 'Explorar Ferramentas'. The main content area is titled 'Dados' and displays a list of public datasets. A blue speech bubble overlay points to the list with the text 'Lista de *datasets* públicos'.

Dataset Name	Id	Criado em	Atualizado em	Ações
Teste - Drebin	4768aeece54-4f5f-b5cf-c1a7e3c1109f	2024-11-25T21:33:49.817Z	2024-11-25T21:35:57.586Z	Baixar dados originais, Gerar dados sintéticos
Teste - Kronodroid	bee18b3f-d8a6-4cd6-8bdc-e0e382633c1f	2024-11-25T21:31:14.793Z	2024-11-25T21:34:26.420Z	Baixar dados originais, Gerar dados sintéticos
Teste - Adroit	1d151c15-9864-4109-9c26-a025a53e5deb	2024-11-25T19:30:56.508Z	2024-11-25T21:29:21.908Z	Baixar dados originais, Gerar dados sintéticos
Android Permissions				

Subir dados



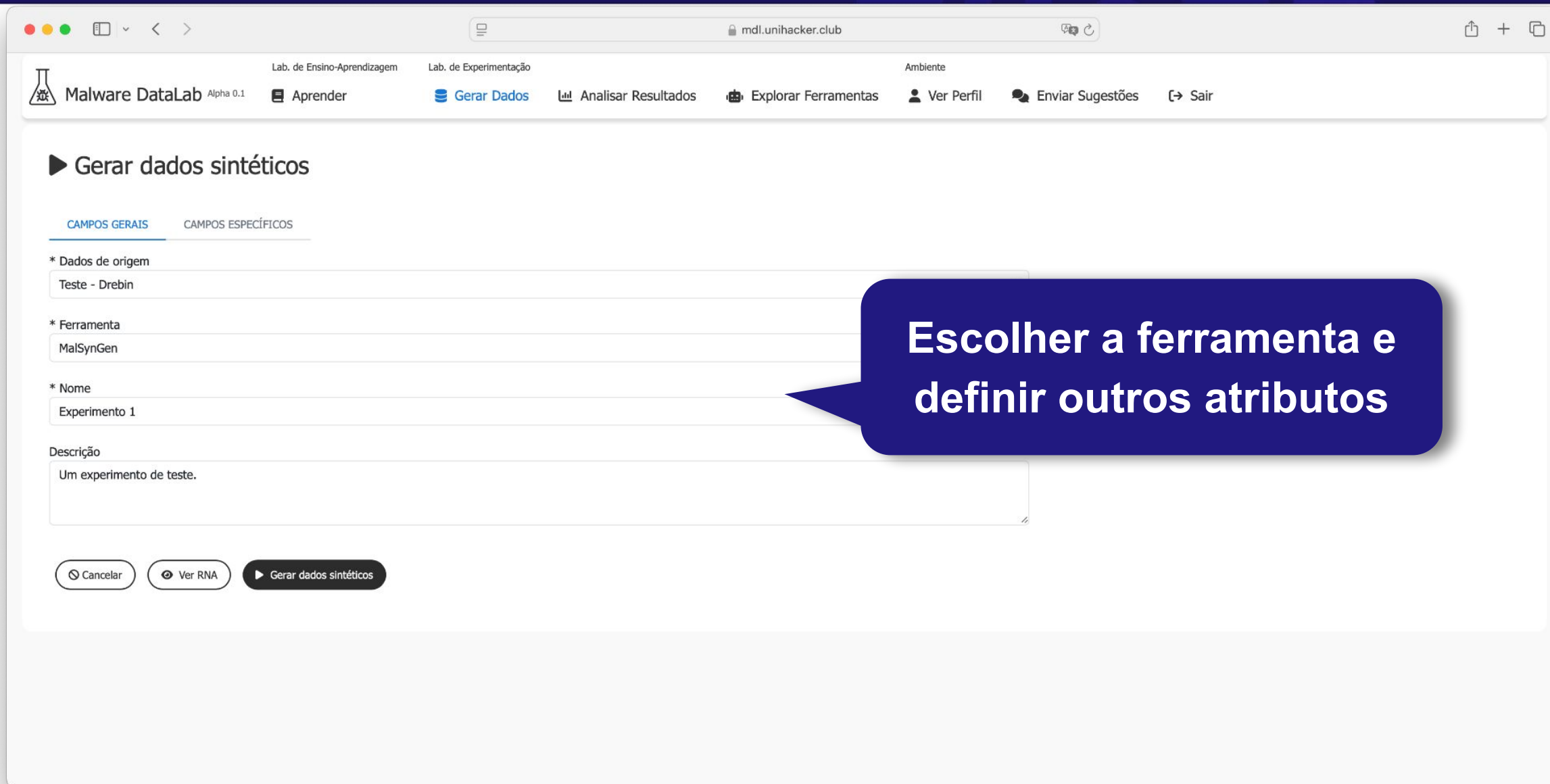
The screenshot shows a web browser window with the URL `mdl.unihacker.club`. The page is titled "Malware DataLab Alpha 0.1" and features a navigation bar with links: "Aprender", "Gerar Dados", "Analisar Resultados", "Explorar Ferramentas", "Ver Perfil", "Enviar Sugestões", and "Sair". The main content area is titled "Novo dataset" and contains three required fields:

- * Nome**: A text input field with a red border and a red error message below it: "O nome é obrigatória".
- * Tags (separadas por ,)**: A text input field containing the text "Exemplo".
- * Arquivo .csv**: A file upload section with a "Choose File" button and the text "no file selected".

At the bottom of the form are two buttons: "Cancelar" and "Enviar".

**Upload de novo *dataset*:
privado ou público**

Gerar dados



The screenshot shows a web browser window with the URL `mdl.unihacker.club`. The page header includes the 'Malware DataLab Alpha 0.1' logo and navigation links: 'Aprender', 'Gerar Dados' (highlighted), 'Analisar Resultados', 'Explorar Ferramentas', 'Ver Perfil', 'Enviar Sugestões', and 'Sair'. The main content area is titled 'Gerar dados sintéticos' and features two tabs: 'CAMPOS GERAIS' (active) and 'CAMPOS ESPECÍFICOS'. The form contains the following fields:

- * Dados de origem:** A dropdown menu with 'Teste - Drebin' selected.
- * Ferramenta:** A dropdown menu with 'MalSynGen' selected.
- * Nome:** A text input field containing 'Experimento 1'.
- Descrição:** A text area containing 'Um experimento de teste.'

At the bottom of the form are three buttons: 'Cancelar', 'Ver RNA', and 'Gerar dados sintéticos' (which has a play icon).

Escolher a ferramenta e definir outros atributos

Ver RNA

Visualizar a RNA da
ferramenta escolhida

mdl.unihacker.club

Malware DataLab Alpha 0.1

Lab. de Ensino-Aprendizagem Lab. de Experimentação Ambiente

Aprender Gerar Dados Analisar Resultados Explorar Ferramentas Ver Perfil Enviar Sugestões Sair

* Ferramenta
MalSyn

* Nome
Experimento 1

Descrição
Um experimento de te

Cancelar Ver RNA Gerar dados sintéticos

Estrutura do discriminador

Dado de entrada Rótulos Real

Camada de Embedding

Blocos Intermediários

Camada Densa

Ativação Sigmoid

Concatenação

Camada Densa

Ativação Intermediária

Dropout

Estrutura do gerador

Ruído Latente Rótulos

Camada de Embedding

Blocos Intermediários

Camada Densa

Ativação Relu

Concatenação

Camada Densa

Ativação Intermediária

Dropout

Imagem real

Rótulos

Ruído

Gerador

Discriminador

Perda

HACKERS
DO
BEM

Iniciado em

***Job* aguardando na fila
para ser executado**

Execução

mdl.unihacker.club

 **Malware DataLab** Alpha 0.1

Lab. de Ensino-Aprendizagem

Lab. de Experimentação

Ambiente

 Aprender

 Gerar Dados

 Analisar Resultados

 Explorar Ferramentas

 Ver Perfil

 Enviar Sugestões

 Sair

Resultados

Pesquisar

 **Aguardando na fila para ser executado**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

 **Em execução**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:01:59.419Z

 **Concluído**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:00:48.144Z

Finalizado em

2024-11-28T21:01:20.879Z

 Baixar dados sintéticos

 Baixar métricas

 Ver métricas

 **Concluído**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:00:39.314Z

Finalizado em

2024-11-28T21:01:12.138Z

 Baixar dados sintéticos

 Baixar métricas

 Ver métricas

Job em execução



29

Execução

mdl.unihacker.club

 **Malware DataLab** Alpha 0.1

Lab. de Ensino-Aprendizagem

Lab. de Experimentação

Ambiente

 Aprender

 Gerar Dados

 Analisar Resultados

 Explorar Ferramentas

 Ver Perfil

 Enviar Sugestões

 Sair

Resultados

Pesquisar

 **Aguardando na fila para ser executado**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

 **Em execução**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:01:59.419Z

 **Concluído**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:00:48.144Z

Finalizado em

2024-11-28T21:01:20.879Z

 Baixar dados sintéticos

 Baixar métricas

 Ver métricas

 **Concluído**

Ferramenta

MalSynGen

Dataset

Teste - Adroit

Iniciado em

2024-11-28T21:00:39.314Z

Finalizado em

2024-11-28T21:01:12.138Z

 Baixar dados sintéticos

 Baixar métricas

 Ver métricas

**Job concluído
(Ver métricas)**



Analisar Resultados

Curvas de Treinamento

Teste - Adroit, sintetizado por MalSynGen

PARÂMETROS CURVAS DE TREINAMENTO MÉTRICAS DE SIMILARIDADE MÉTRICAS DE APLICABILIDADE MÉTRICAS DE CONFIABILIDADE

Interação entre Gerador e Discriminador

As figuras ilustram a interação entre o gerador e o discriminador durante o treinamento.

Gerador: tem o objetivo de criar amostras que sejam indistinguíveis das reais.

Discriminador: busca melhorar continuamente sua capacidade de distinguir entre amostras reais e falsas.

Competição e Convergência: essa competição entre gerador e discriminador resulta em uma evolução constante das amostras geradas.

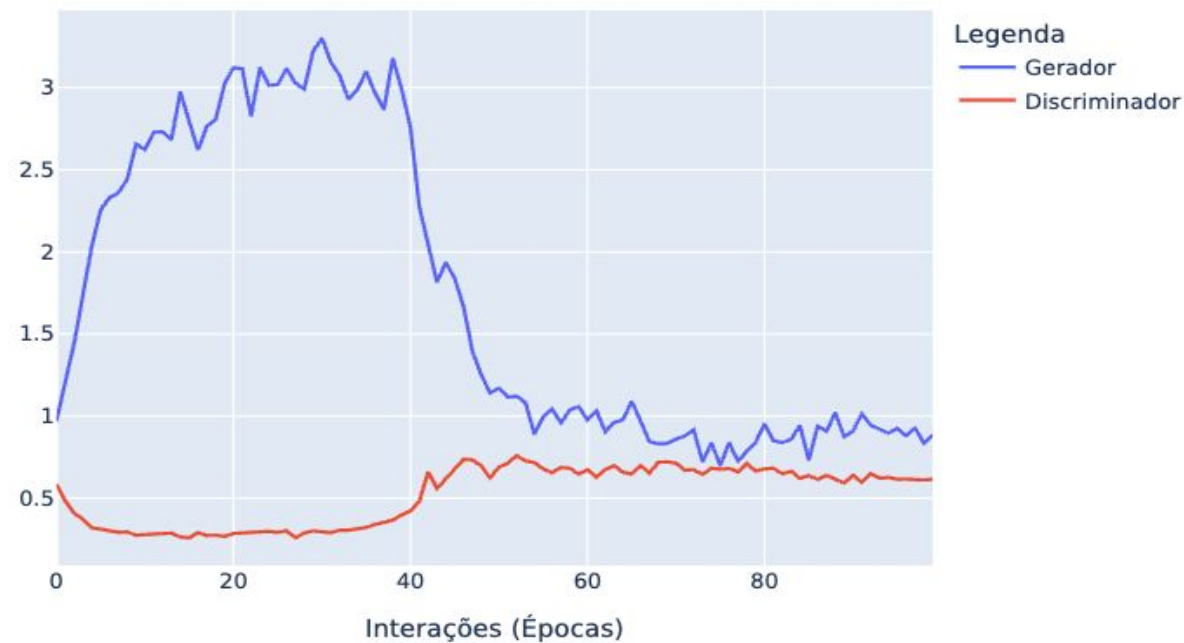
A não convergência das redes GAN pode ser identificada monitorando as curvas de perda. Estas devem apresentar uma tendência de estabilização.

Para cada fold (dobra) de treinamento, é necessário analisar o desempenho.

Perda do Gerador e Discriminador



Perda do Gerador e Discriminador



Analisar Resultados

Métricas de Similaridade

Teste - Adroit, sintetizado por MalSynGen

PARÂMETROS CURVAS DE TREINAMENTO MÉTRICAS DE SIMILARIDADE MÉTRICAS DE APLICABILIDADE MÉTRICAS DE CONFUSÃO

As métricas de similaridade avaliam se os dados gerados diferem dos dados originais,

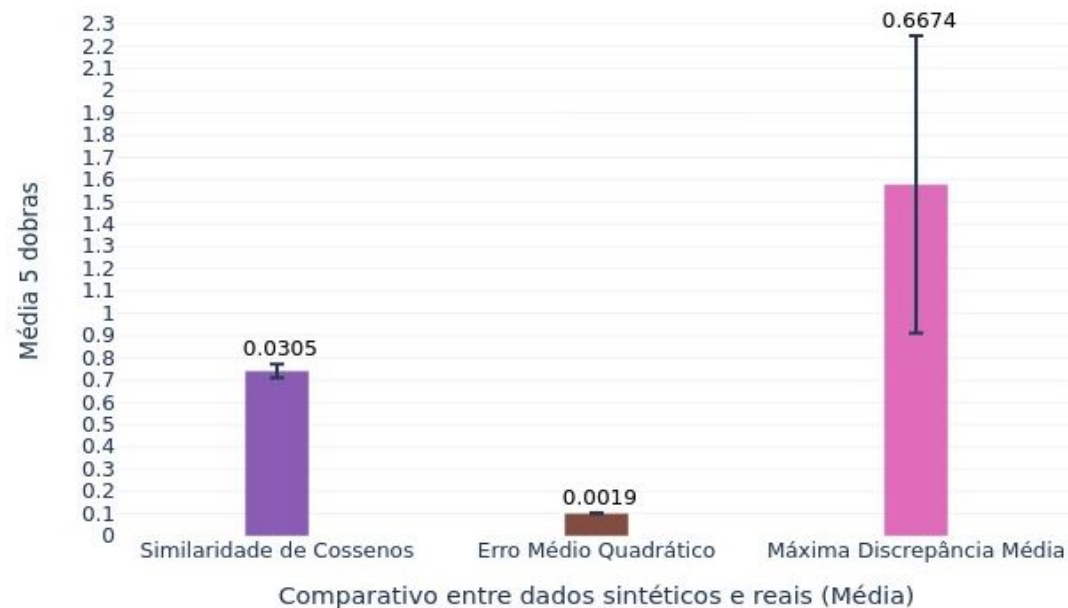
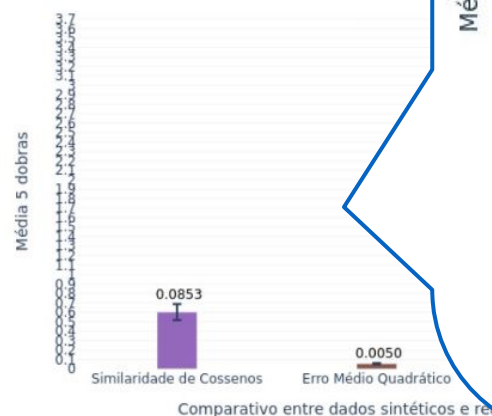
Similaridade de Cossenos: um valor próximo de 1 (ou se aproximando, mas não exatame

Erro Médio Quadrático: um valor próximo de zero indica m

Máxima Discrepância Média: um valor próximo de zero indica

A análise é feita para amostras m

Amostras Ma



Amostras Benignas

Analisar Resultados

Métricas de Aplicabilidade

Teste - Adroit, sintetizado por MalSynGen

PARÂMETROS CURVAS DE TREINAMENTO MÉTRICAS DE SIMILARIDADE MÉTRICAS DE APLICABILIDADE MÉTRICAS DE CONFUSÃO

As Métricas de Aplicabilidade avaliam se os classificadores conseguem classificar os dados sintéticos de forma similar.

Abordagens Utilizadas

TR-AS (Train on Real, Test on Synthetic - TRAs): o modelo é Treinado com dados reais e avaliado com dados sintéticos.

TS-AR (Train on Synthetic, Test on Real - TSAR): o modelo é Treinado com dados sintéticos e avaliado com dados reais.

Para cada abordagem é apresentado um gráfico para cada modelo.

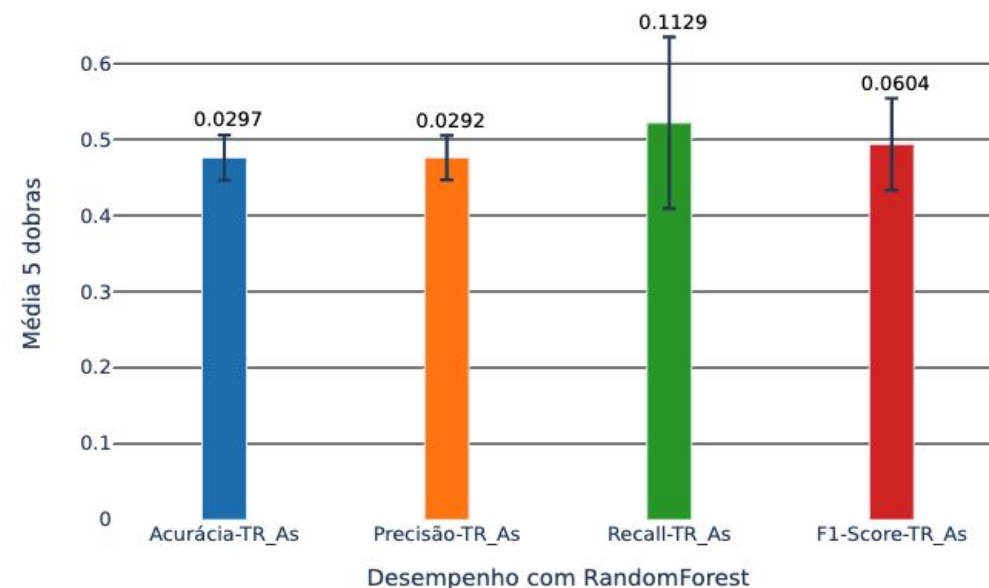
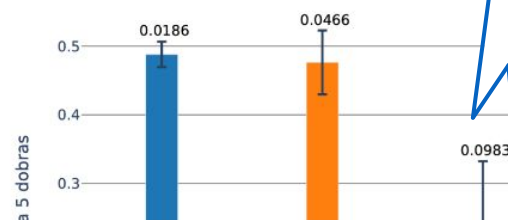
Interpretação dos Resultados

Para métricas de classificação binária:

Quanto mais próximo de 1, melhor é o desempenho do classificador ao utilizar os dados sintéticos. A métrica **recall** é considerada mais importante que a **precisão**.

RANDOMFOREST SUPPORTVECTORMACHINE

TR As: Treinado com Real e Avaliado com Sintético



TS Ar: Treinado com Sintético e Avaliado com Real

Analisar Resultados

Matrizes de Confusão

Teste - Adroit, sintetizado por MalSynGen

PARÂMETROS CURVAS DE TREINAMENTO MÉTRICAS DE SIMILARIDADE MÉTRICAS DE APLICABILIDADE MÉTRICAS DE CONFUSÃO

Em geral, as matrizes de confusão são usadas para avaliar os algoritmos de classificação considerando algum dado de entrada. Nós usamos para v com dados sintéticos se assemelham aos resultados obtidos com os dados reais, m

As matrizes de confusão permitem identificar os acertos e erros do

Falsos Positivos (FP) Falsos Negativos (FN) Verdadeiros Positivos (V

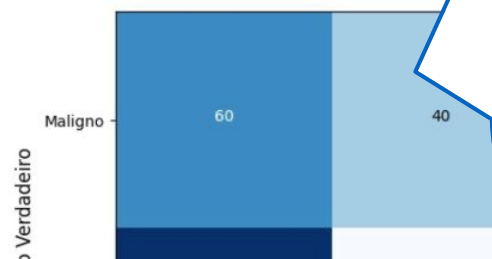
Em cenários específicos, como a classificação entre "benigno" e "maligno", a performance do modelo é avalia

Maligno-Maligno (Verdadeiros Positivos) (quadrante superior esquerdo) **Benigno-Benigno** (Verdadeiros Negativos) (quadrante inferior direito)
Quanto maiores forem esses valores (visualmente, tons mais escuros na diagonal principal), e consequentemente menor os outros valores (visualmente, tons mais claros na diagonal secundária), melhor o modelo realiza classificações corretas e confiáveis

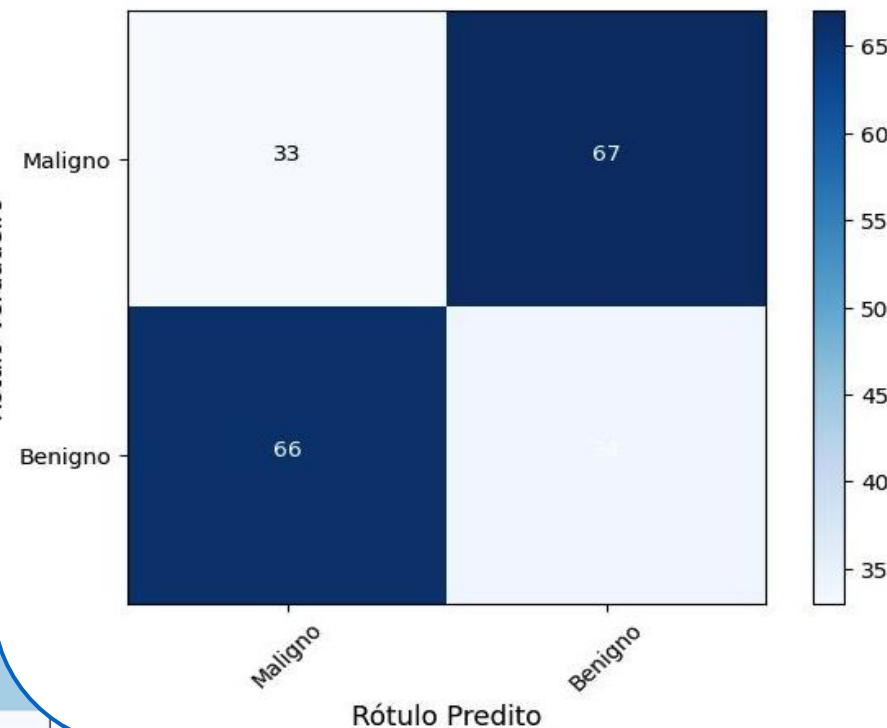
RANDOMFOREST SUPPORTVECTORMACHINE

TR As: Treinado com Real e Avaliado

Fold 1



Fold 1



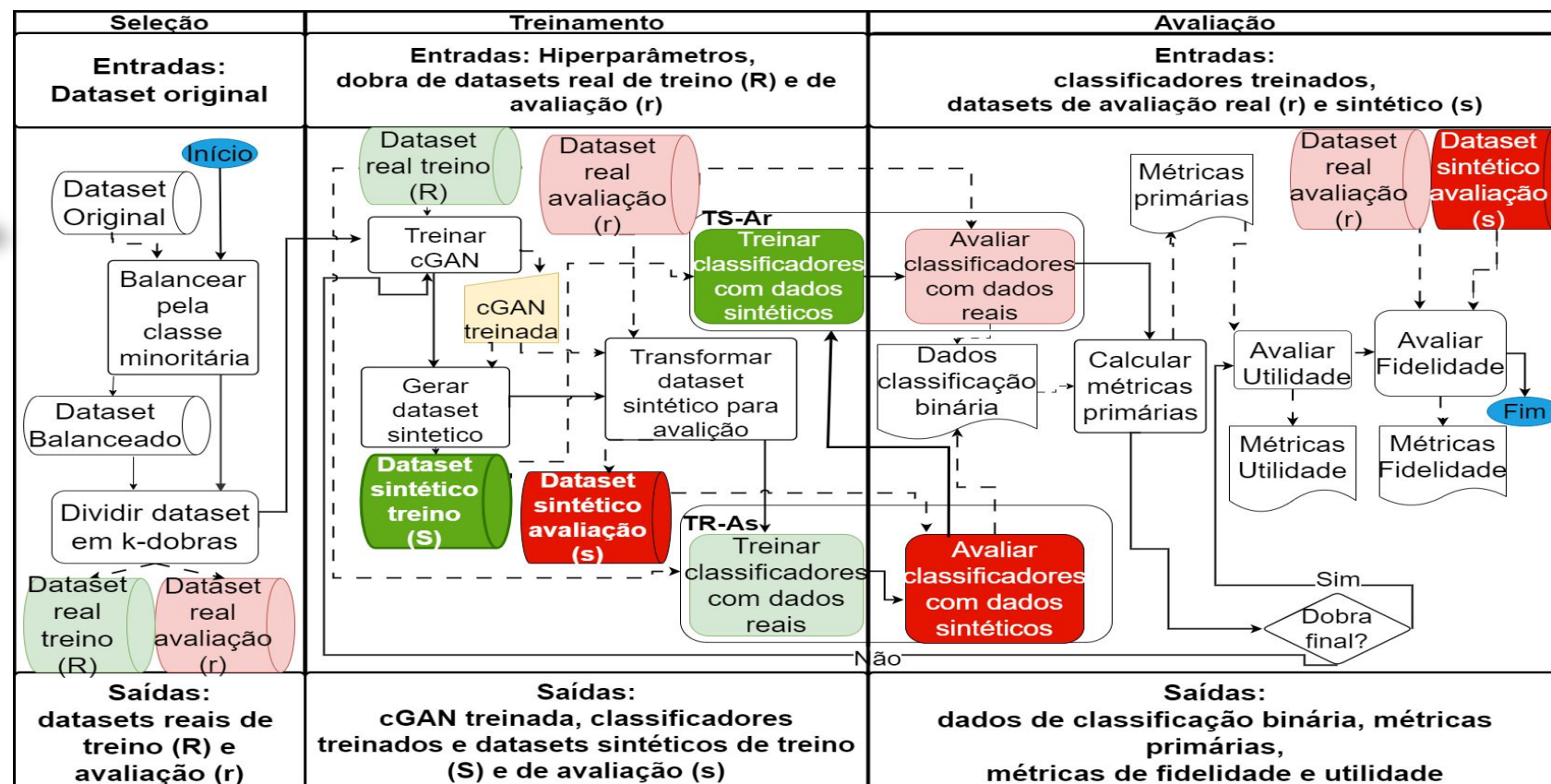
Demo Day Malware DataLab



1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. **MalSynGen (evolução da DroidAugmentor)**
4. Cloud AutoDroid (evolução da AutoDroid)
5. GUI4MalSynGen (versão minimalista)

MalSynGen

Nova arquitetura e
nova metodologia
de avaliação



MalSynGen: redes neurais artificiais na geração de dados tabulares sintéticos para detecção de malware

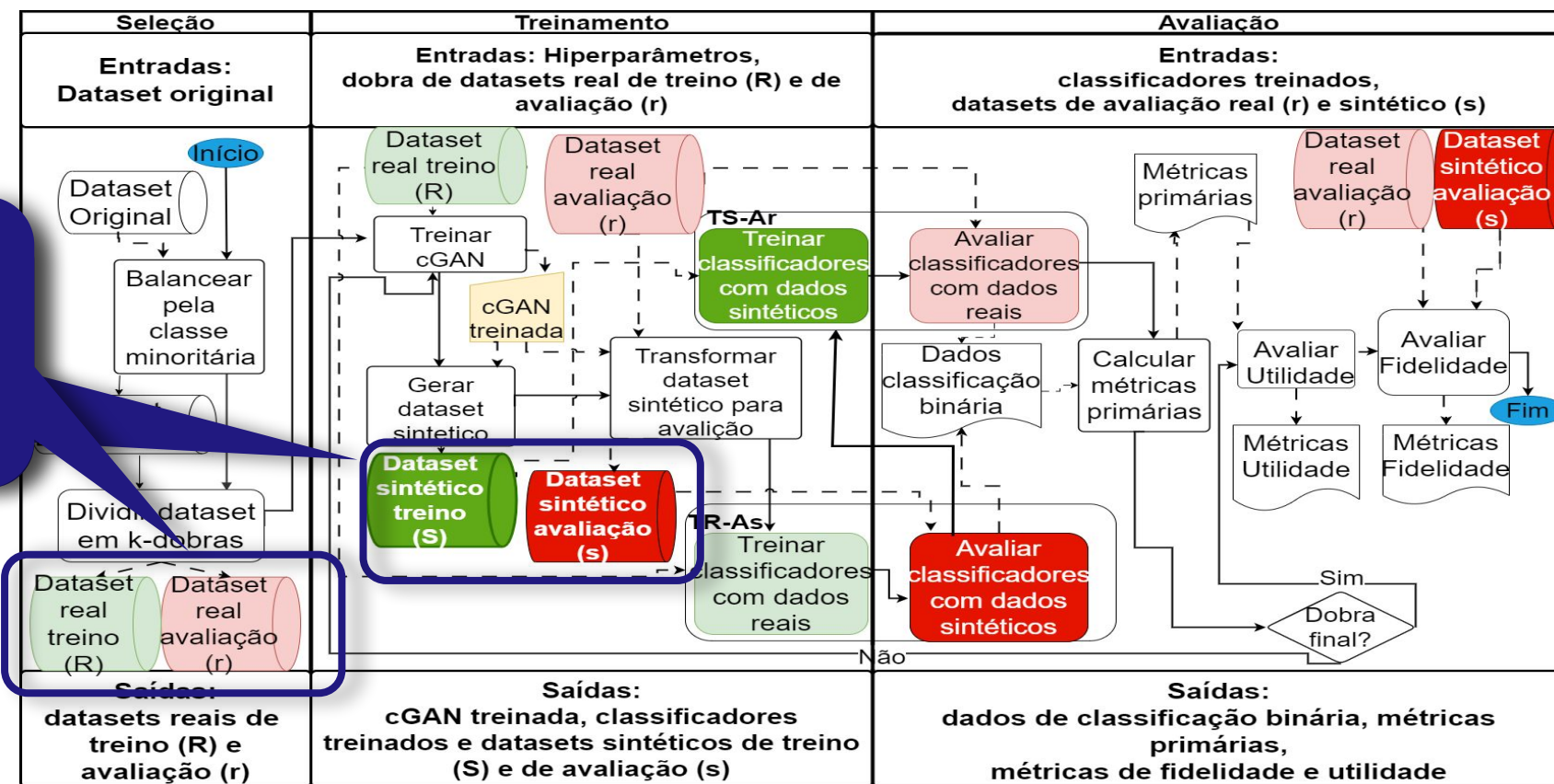
Angelo Gaspar, Diego Kreutz, Hendrio Bagança, Rodrigo Mansilha, Kayuã Oleques Paim

VIII Salão de Ferramentas do SBSEG 2024

https://sol.sbc.org.br/index.php/sbseg_estendido/article/view/30126

MalSynGen

Datasets (real e sintético) de treino e avaliação



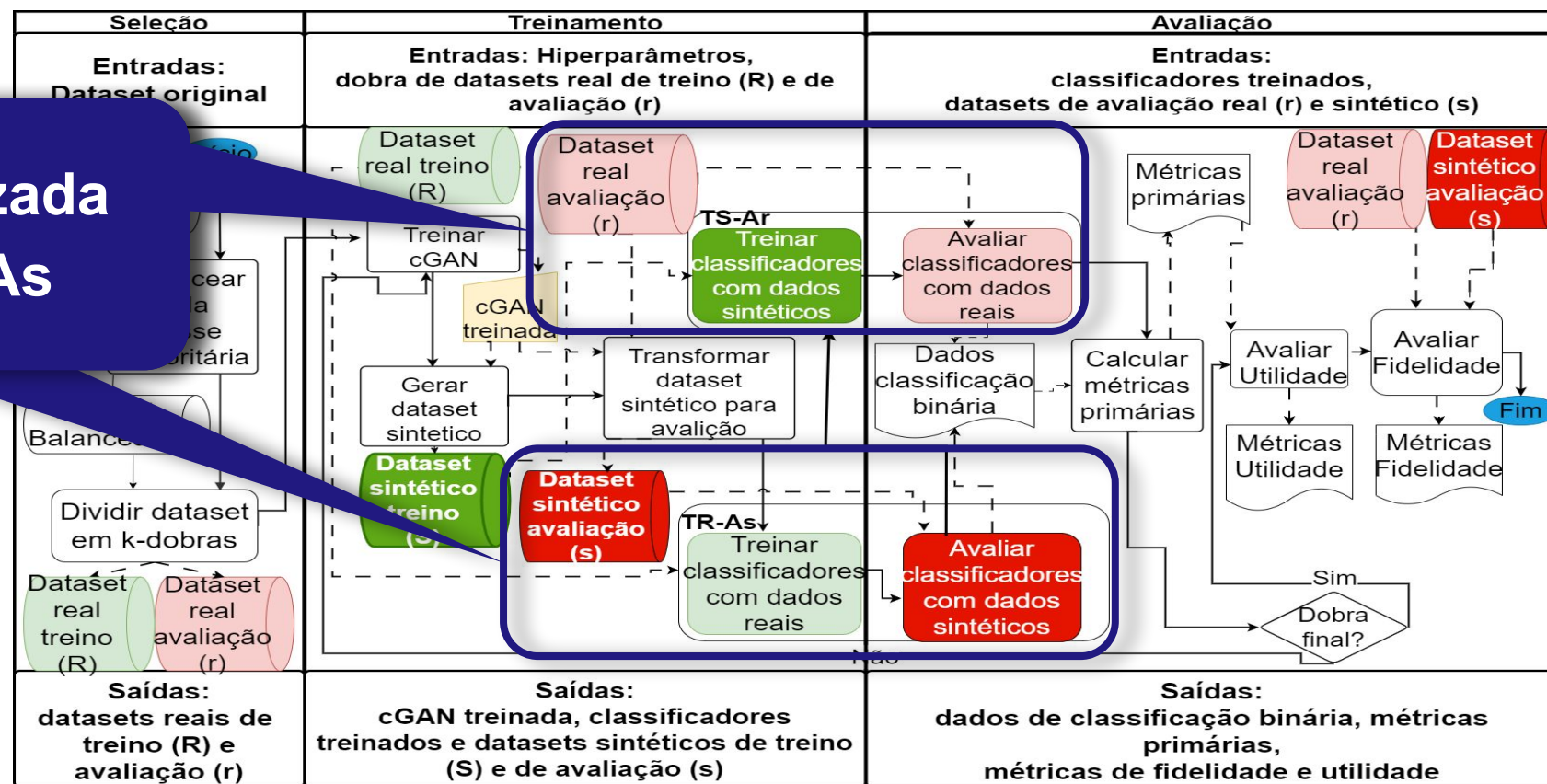
MalSynGen: redes neurais artificiais na geração de dados tabulares sintéticos para detecção de malware

Angelo Gaspar, Diego Kreutz, Hendrio Bagança, Rodrigo Mansilha, Kayuã Oleques Paim

VIII Salão de Ferramentas do SBSEG 2024

https://sol.sbc.org.br/index.php/sbseg_estendido/article/view/30126

Avaliação cruzada TS-Ar e TR-As

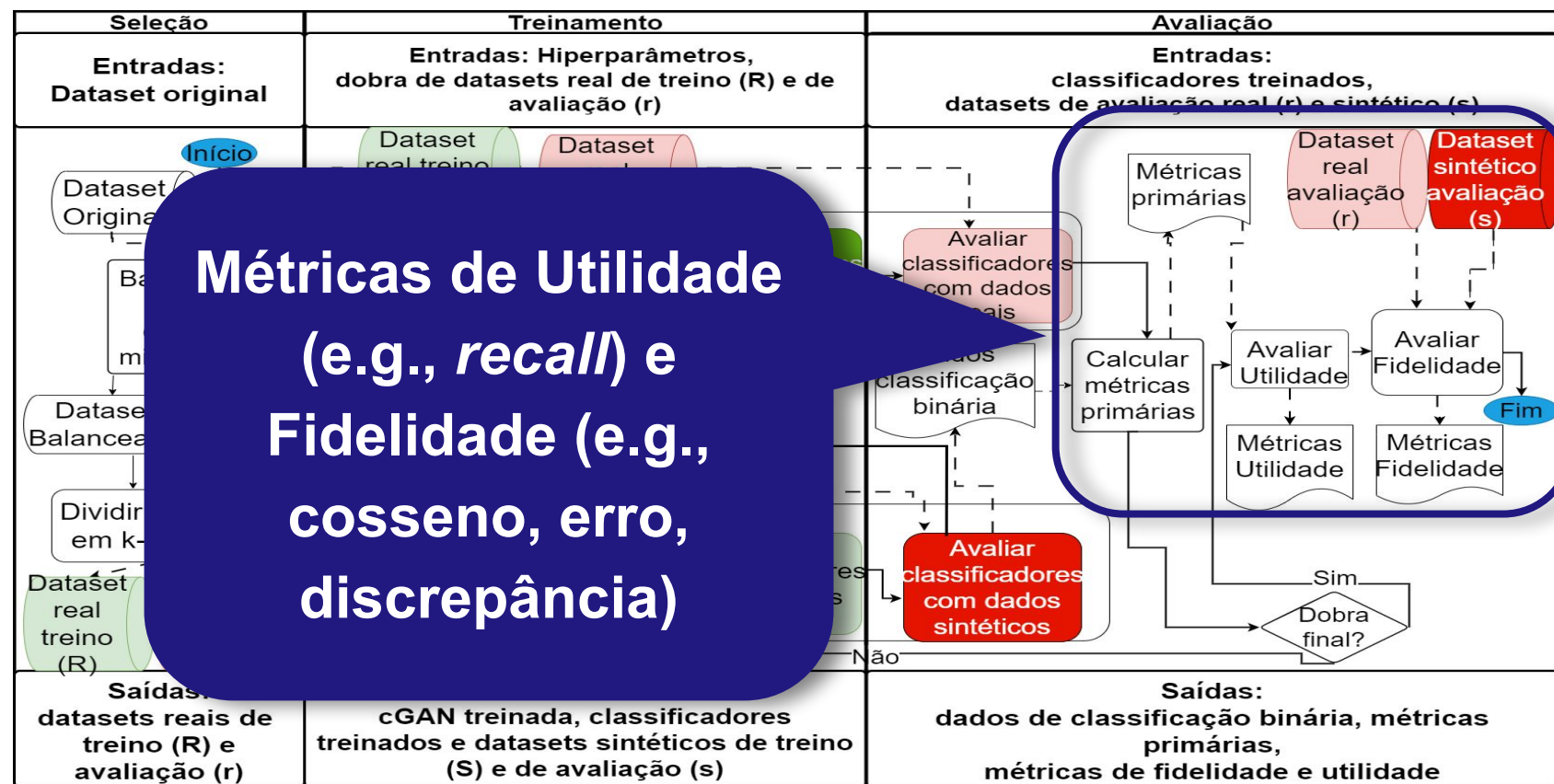


MalSynGen: redes neurais artificiais na geração de dados tabulares sintéticos para detecção de malware

Angelo Gaspar, Diego Kreutz, Hendrio Bagança, Rodrigo Mansilha, Kayuã Oleques Paim

VIII Salão de Ferramentas do SBSEG 2024

https://sol.sbc.org.br/index.php/sbseg_estendido/article/view/30126



Métricas de Utilidade
(e.g., *recall*) e
Fidelidade (e.g.,
cosseno, erro,
discrepância)



MalSynGen: redes neurais artificiais na geração de dados tabulares sintéticos para detecção de malware

Angelo Gaspar, Diego Kreutz, Hendrio Bagança, Rodrigo Mansilha, Kayuã Oleques Paim

VIII Salão de Ferramentas do SBSEG 2024

https://sol.sbc.org.br/index.php/sbseg_estendido/article/view/30126

Título: Enabling Data-driven Innovation with Synthetic Data



Vyas Sekar

Carnegie Mellon University

Resumo: Today in computer systems and security research, lack of access to realistic and diverse data from multiple deployments hampers innovation; e.g., products trained on data not representative of environment, there is no way to quantitatively assess products; machine learning workflows experiences data drift, and product audit/feedback is not quantitative. The result today is poor in debugging/reproduction/resolution, and impossibility to share it.

In this talk, we will discuss our research outcomes on demonstrating synthetic data using Deep Generative Models (DGMs) to address telemetry, anomaly detection, model training). We have identified privacy challenges and tradeoffs in existing approaches. By synthesizing advances in machine learning and privacy, we identify design choices that will present some of the key results from our work in applying these to relevant datasets and use cases.



Mini Bio: Vyas Sekar is the Tan Family Professor of Electrical and Computer Engineering in the ECE Department at Carnegie Mellon University, a Senior Technologist at Rockfish Data, and the Chief Scientist at Conviva.

His research is broadly at the intersection of networks, systems, and security. His work has been recognized with the ACM SIGCOMM Test of Time Award, the Science of Security prize, the Intel Outstanding Researcher Award, the SIGCOMM Test of Time Award, and the

Demo Day Malware DataLab

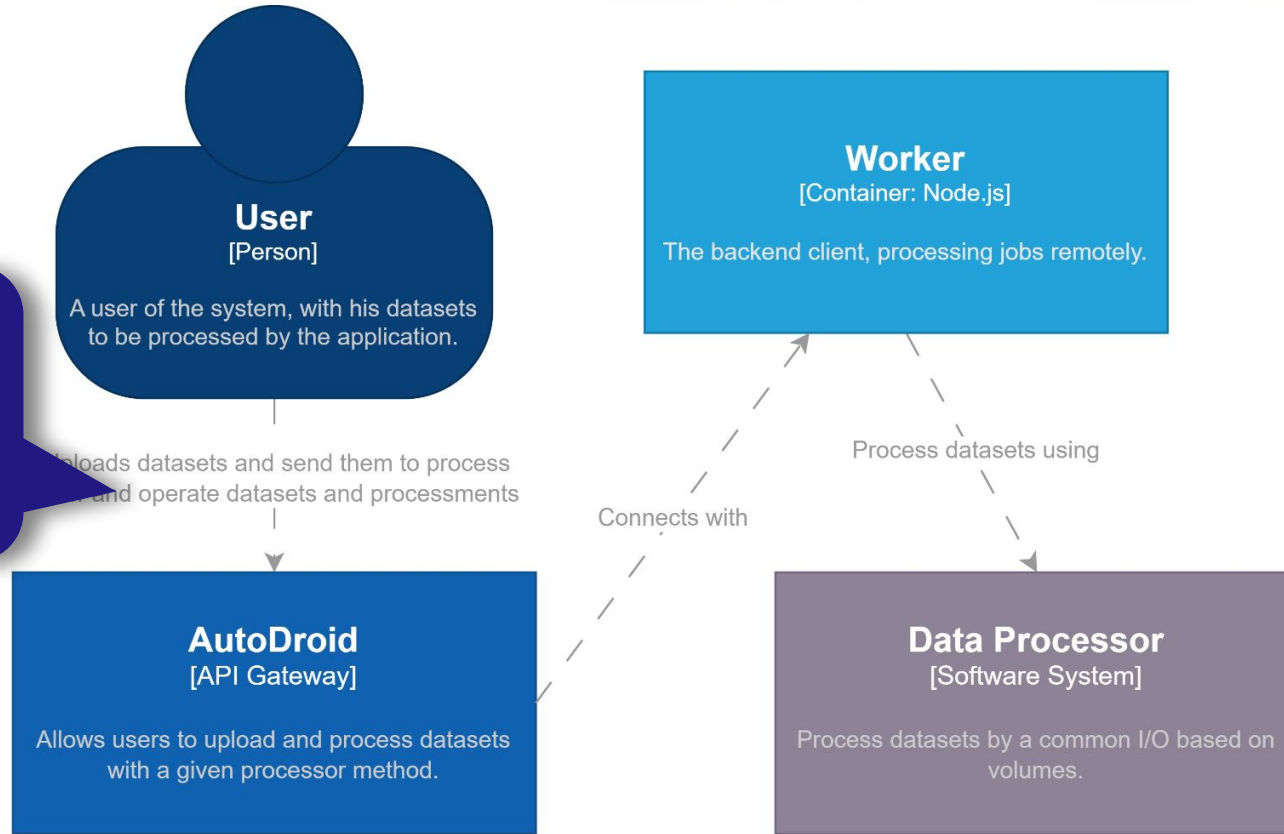


1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. MalSynGen (evolução da DroidAugmentor)
4. **Cloud AutoDroid (evolução da AutoDroid)**
5. GUI4MalSynGen (versão minimalista)

Cloud AutoDroid: sistema distribuído



API (REST e GraphQL)
Workers distribuídos



Legend

Person
Software System
Container
Component
External Person
External Software System



(paper)

[System Context] AutoDroid

The system context diagram for the Data Processor Gateway System.

Cloud AutoDroid: uma Arquitetura de Backend para Executar Serviços de IA Generativa na Nuvem

Luiz Felipe Laviola, Angelo Gaspar Diniz Nogueira, Diego Kreutz, Rodrigo Brandão Mansilha

VIII Escola Regional de Engenharia de Software (ERES 2024)

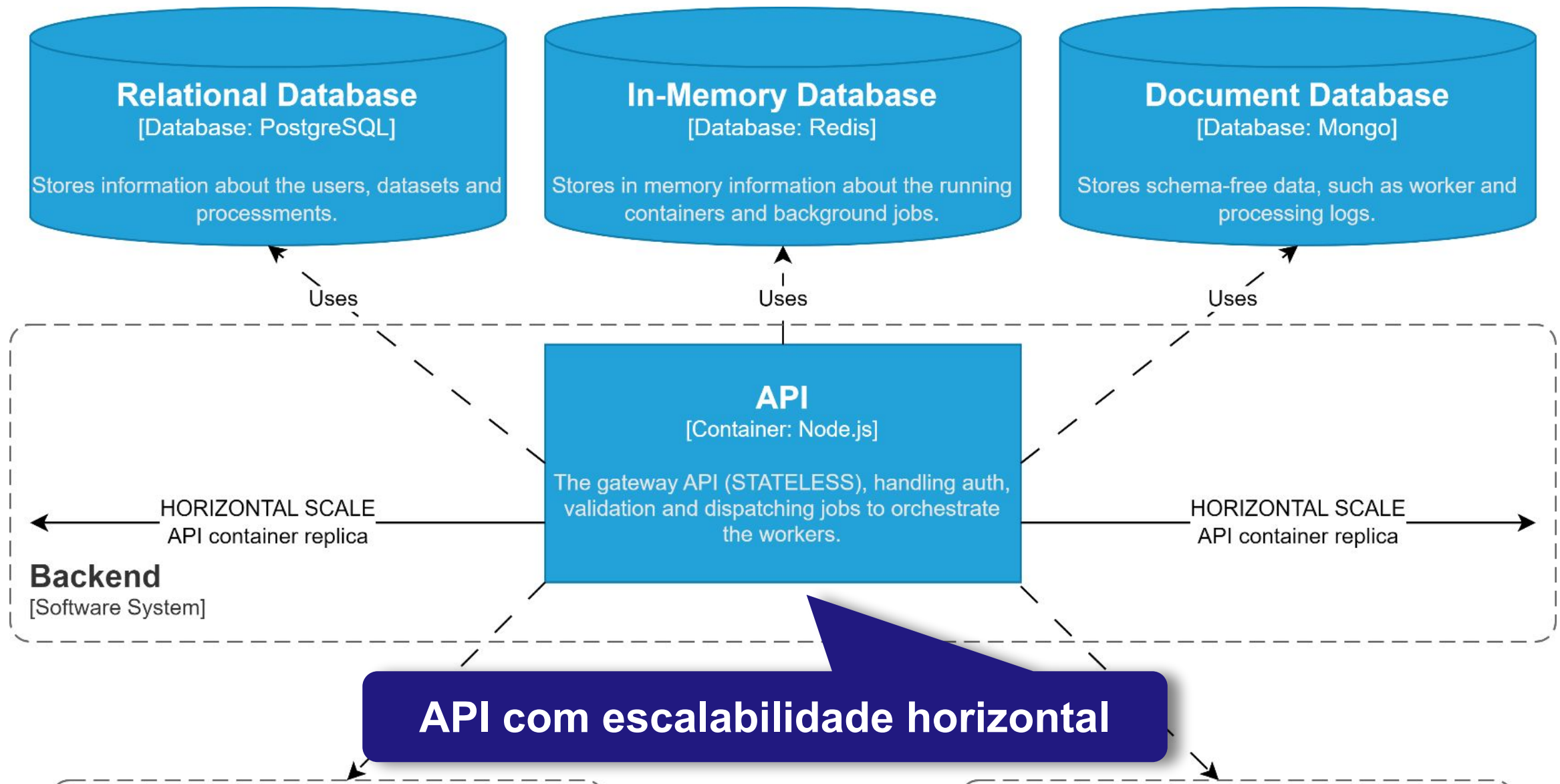
<https://drive.google.com/file/d/1Q8acpITbFMp9cVdsMpPvpJghMf4ITQqH/view>



(slides)

42

Cloud AutoDroid: sistema distribuído



Cloud AutoDroid: limitações atuais



- **Infraestrutura atual:**
 - backend e frontend: 2 máquinas
 - workers (máquinas de processamento): 6 disponíveis
- **Armazenamento: cota gratuita** do Firebase Storage limitado a **5GB** e largura de banda limitada e **1GB** por dia
- **Auto-escalabilidade:** necessita recursos de cloud e implementação de auto-deploy de workers

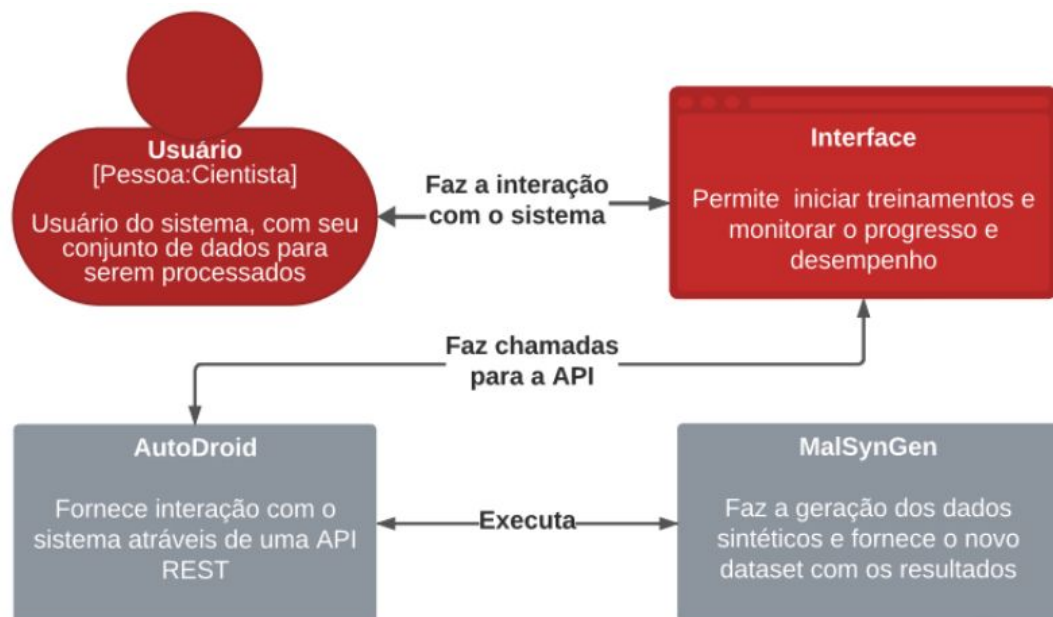


Demo Day Malware DataLab



1. Laboratório de Ensino-Aprendizagem
2. Laboratório de Experimentação
3. MalSynGen (evolução da DroidAugmentor)
4. Cloud AutoDroid (evolução da AutoDroid)
5. **GUI4MalSynGen (versão minimalista)**

GUI4MalSynGen



Parâmetros de Treinamento

Campanhas Disponíveis: Campanha ▾

Carregue seu conjunto de dados

Escolher arquivo: Nenhum... escolhido

Tipos de extensão: .csv

Adicionar



Campanhas Seleccionadas

Dense64_E1000

Iniciar



(paper)

**Instância *standalone* para
download e teste local**

Uma GUI para Hackers do Bem Aprenderem Sobre Malwares Sintéticos
Leonardo Karling Sonco, Angelo Nogueira, Diego Kreutz, Rodrigo Brandão Mansilha
VIII Escola Regional de Engenharia de Software (ERES 2024)

<https://github.com/MalwareDataLab/GUI4MalSynGen>



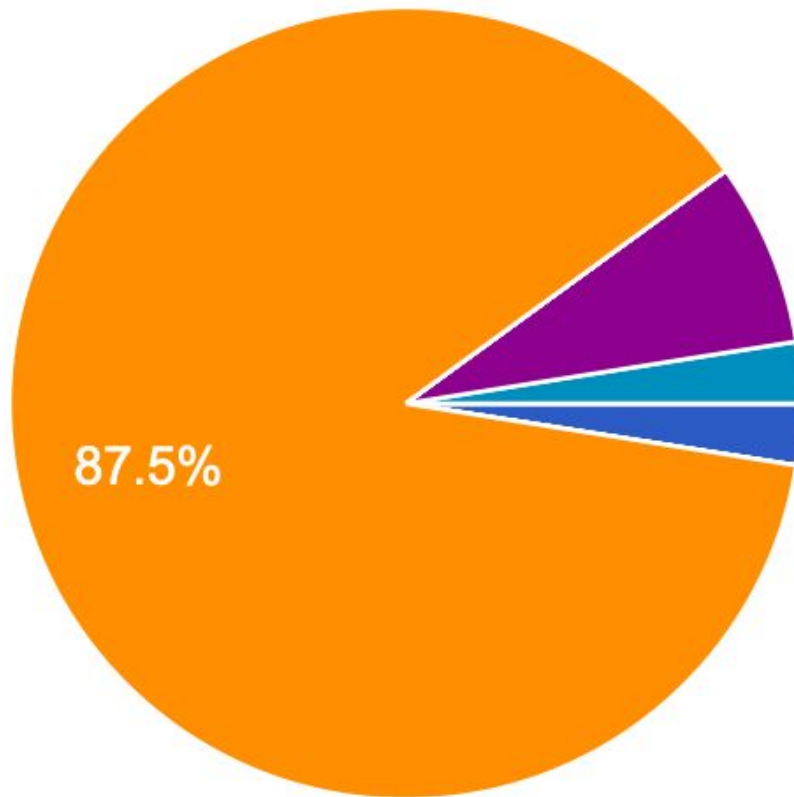
(GitHub)

1. Testes com Usuários
2. Prospeção de Mercado
3. Publicações
4. Premiações
5. Direções futuras

27 de novembro de 2024:

- 59 usuários registrados
- 229 experimentos executados

Testes com Usuários



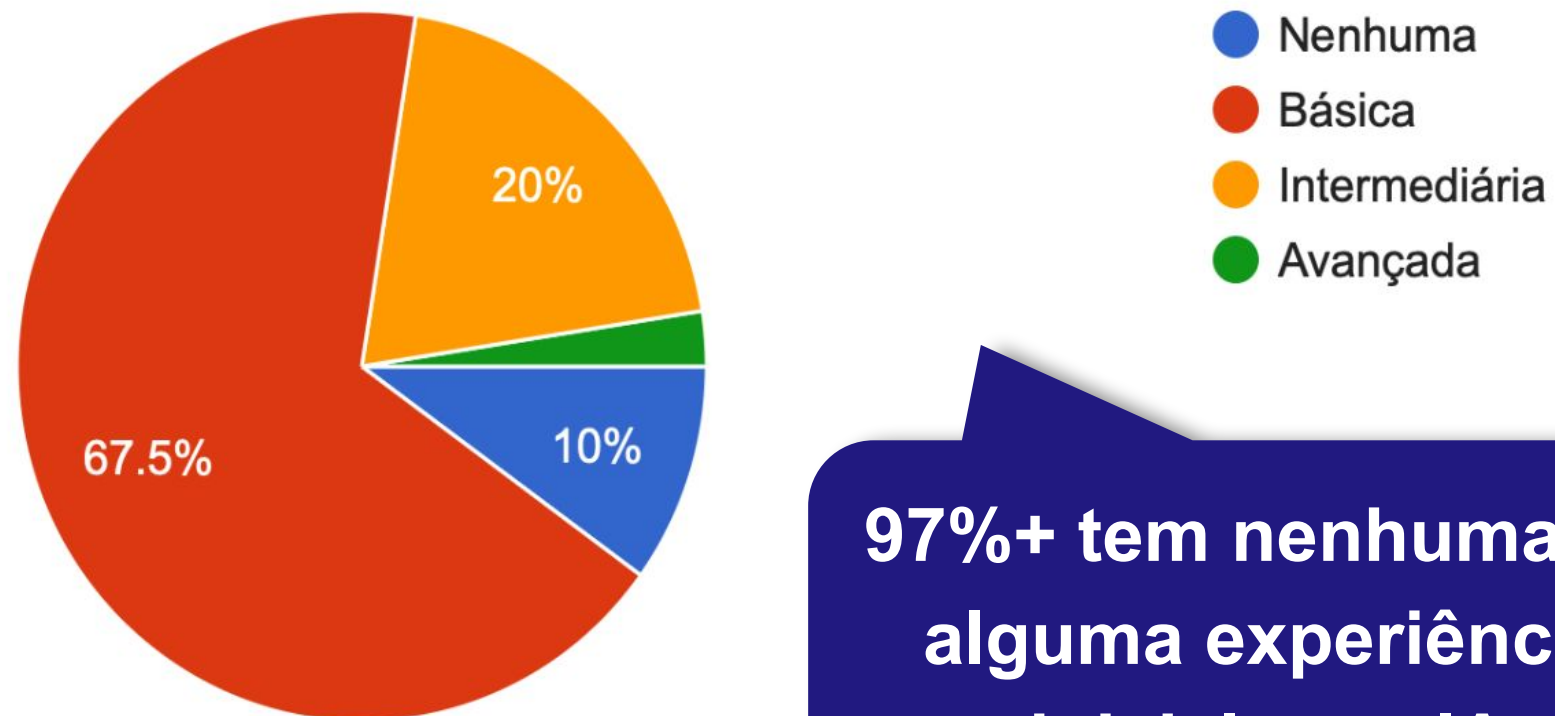
- Ensino Médio
- Ensino Técnico
- Graduação (Bacharelado/Licenciatura)
- Pós-Graduação (Especialização)
- Mestrado
- Doutorado
- Profissional (Atuação no mercado)

4 perfis
40 participantes
4 instituições

Testes com Usuários

Qual é a sua experiência atual com Inteligência Artificial?

40 responses



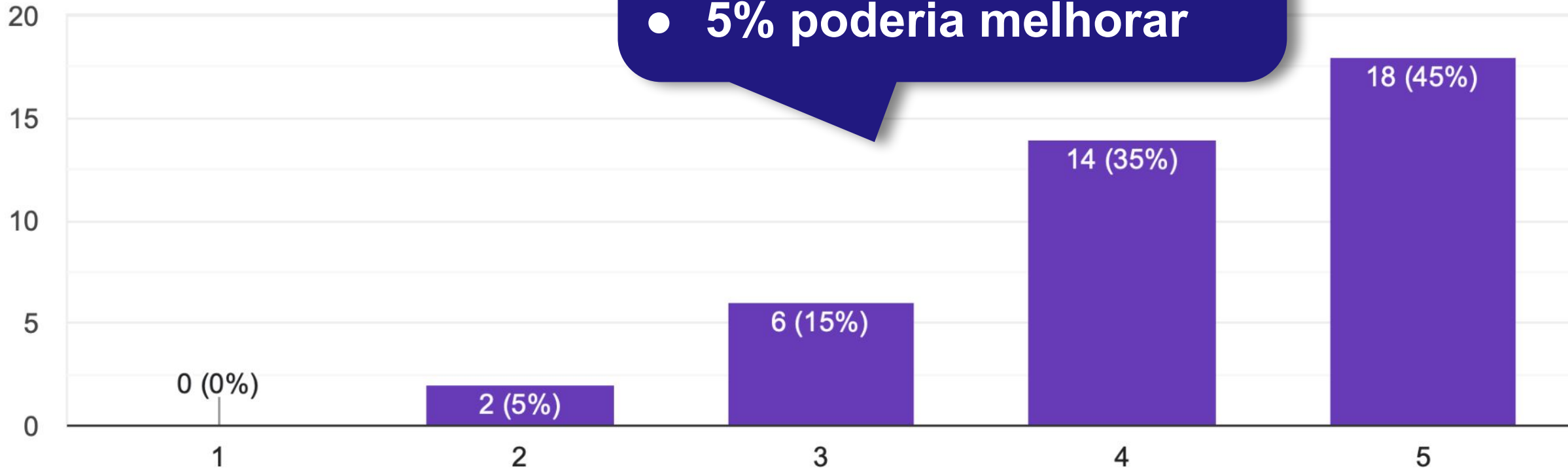
97%+ tem nenhuma ou alguma experiência inicial com IA

Testes com Usuários

É fácil aprender a usar a interface gráfica

40 responses

- 80% é fácil
- 15% intermediário
- 5% poderia melhorar



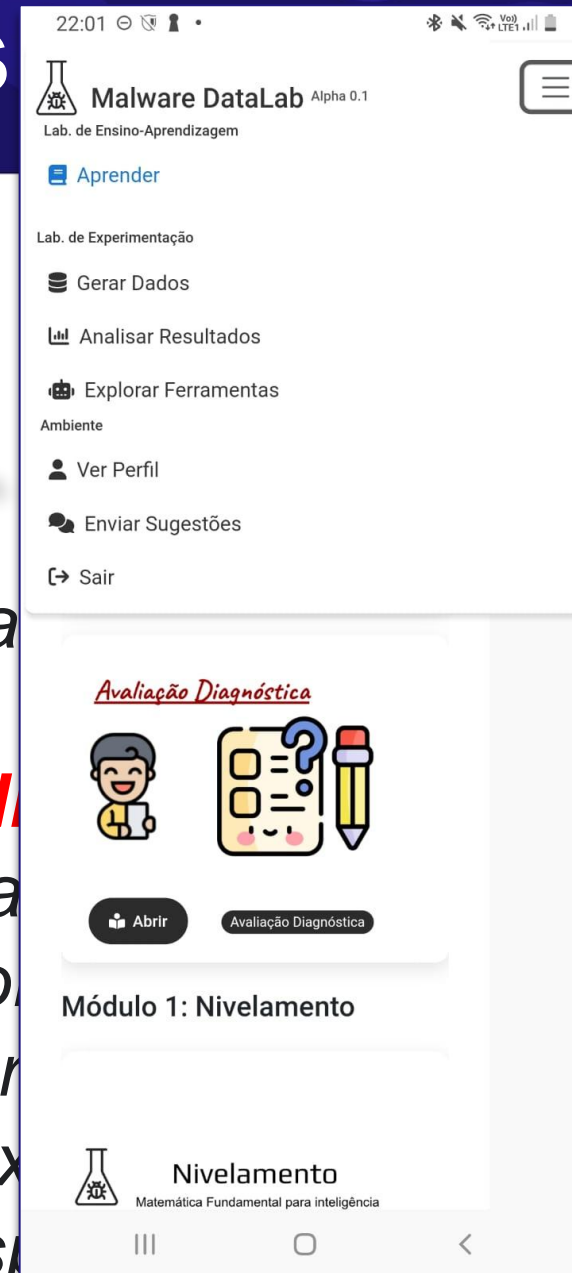
Interface precisa melhorar para versão mobile

"Sobre a questão do uso da plataforma, em alguns modelos de tela dependendo da resolução, os menus não ficam disponíveis da maneira correta. **Para utilização mobile, o menu não é apresentado corretamente**, apenas a parte dos módulos são listados, a pesquisa funciona, porém se eu precisar seguir o passo a passo dentro do mobile para acessar por exemplo: Gerar Dados, Analisar Resultados, Explorar algoritmo e perfil não são acessíveis por dispositivos menores."

Testes com Usuários

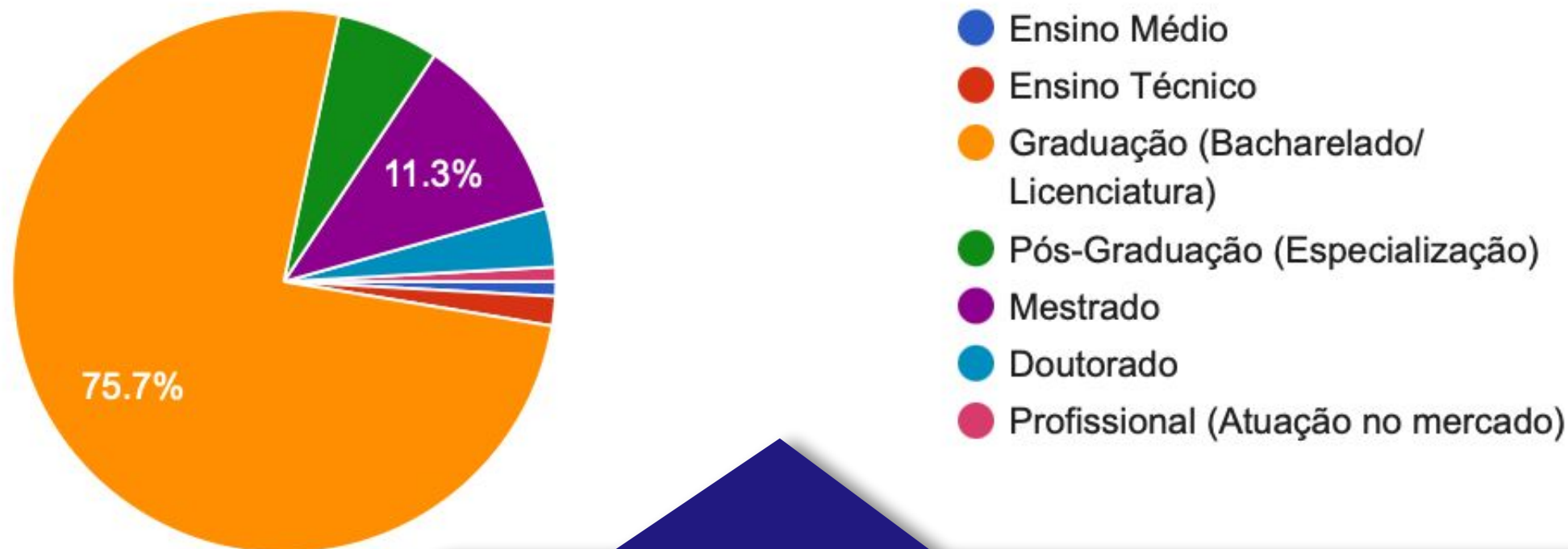
Já estamos
trabalhando nisso!

"Sobre a questão do uso da plataforma dependendo da resolução, da maneira correta. **Para utilizar, o menu não é apresentado corretamente,** a lista de módulos não é apresentada corretamente, a pesquisa funciona, por exemplo, a pesquisa dentro do mobile para gerar dados, analisar resultados, explorar ferramentas, ver perfil, enviar sugestões, sair, etc. são acessíveis por dispositivos."



ma modelos de
am disponíveis
menu não é
s módulos são
r seguir o passo
emplo: Gerar
e perfil não são
s."

Prospecção de Mercado

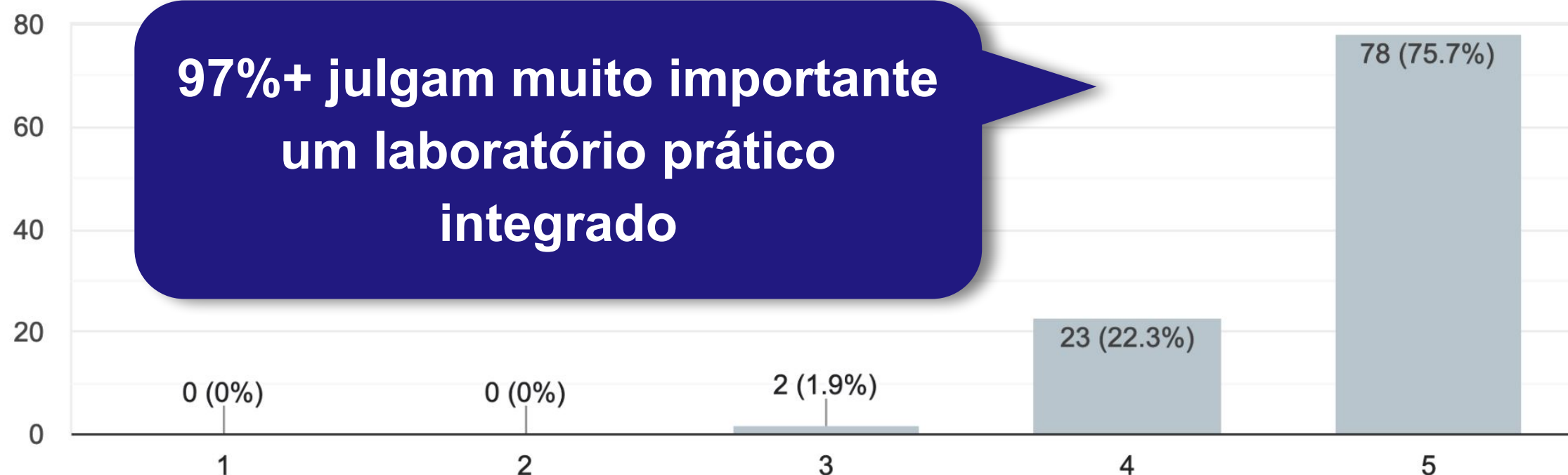


**115 participantes de sul a norte do Brasil
96% tem interesse em formação em IA
aplicada a cibersegurança**

Prospecção de Mercado

Na sua opinião, qual a importância do laboratório prático integrado ao conteúdo de aprendizagem para aprofundar seus conhecimentos e colocar em prática conceitos inovadores de IA?

103 responses



Publicações



Publicações em eventos em 2024: 7

- 32nd Euromicro (internacional): 1
- XXIV SBSEG (nacional): 2
- VIII ERES (regional): 3
- VIII WRSEG (regional): 1

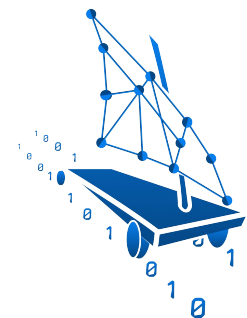


PDP 2024

32nd Euromicro International Conference on
Parallel, Distributed and Network-based Processing



SBSEG24
SÃO JOSÉ DOS CAMPOS



**ERRC
2024**
RIO GRANDE - RS



VIII ESCOLA REGIONAL DE ENGENHARIA DE SOFTWARE
Santiago/RS 11 A 13 DE NOVEMBRO

Capítulos de livro / minicursos em 2024: 2

1. DinD-Bench: Impacto de Contêineres Docker em Docker para a Programação Paralela (ERAD)

Claudio Schepke, Felipe Fava, Diego Kreutz

DOI: <https://doi.org/10.5753/sbc.13779.0.3>



2. Uma Introdução sobre Redes Adversárias Generativas (GAN) e suas Aplicações na Cibersegurança (capítulo em ebook de Cibersegurança do UniHacker)

Karina, Angelo, Anna, Kayua, Diego, Rodrigo, Hendrio



UNIHACKER.CLUB
PROGRAMA UNIVERSIDADE HACKER





Certificamos que o trabalho **“Geração de dados sintéticos tabulares para detecção de Malwares Android: um estudo de caso”**, de autoria de Angelo Diniz (UNIPAMPA), Kayuã Paim (UFRGS), Rodrigo Mansilha (UNIPAMPA) e Diego Kreutz (UNIPAMPA) recebeu o prêmio de

MENÇÃO HONROSA ARTIGO CURTO

na Trilha Principal - Sistemas do 24º Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais.

Lourenço Alves Pereira Jr., Ph.D
Coordenador Geral
Instituto Tecnológico de Aeronáutica

Diego Kreutz, Ph.D
Coordenador Geral
Universidade Federal do Pampa

**Trilha
Principal -
Sistemas**



SBSeg24

SÃO JOSÉ DOS CAMPOS



Certificamos que o trabalho **“MalSynGen: redes neurais artificiais na geração de dados tabulares sintéticos para detecção de malware”**, de autoria de Angelo Gaspar Diniz Nogueira (UNIPAMPA), Kayua Oleques Paim (UFAM), Rodrigo Mansilha (UNIPAMPA), Hendrio Luis Bragança (UFRGS) e Diego Kreutz (UNIPAMPA) recebeu o prêmio de

ARTEFATO DESTAQUE

no Salão de Ferramentas (SF) do 24º Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais.

Lourenço Alves Pereira Jr., Ph.D
Coordenador Geral
Instituto Tecnológico de Aeronáutica

Diego Kreutz, Ph.D
Coordenador Geral
Universidade Federal do Pampa



Premiações (3)



Revisores Destaque:

- **SBSeg24 / CTA:** Angelo Diniz, Diego Kreutz
- **SBSeg24 / WTICG:** Diego Kreutz, Rodrigo Mansilha
- **SBSeg24 / SF:** Diego Kreutz, Rodrigo Mansilha

Concorrendo ao *Best Paper* do WRSeg 2024

Otimização de hiperparâmetros da DroidAugmentor para geração de dados sintéticos de malware Android

Angelo Nogueira, Lucas Oliveira, Anna Luiza Gomes, Diego Kreutz, Rodrigo Mansilha

VIII Workshop Regional de Segurança da Informação e de Sistemas

Computacionais / ERRC 2024



Direções futuras: Interface

Planejado para Dezembro de 2024:

- **Ajustes finos na interface** a partir do feedback dos usuários
- **Análise de UI e UX** com outros colaboradores experientes
 - **Dr. Williamson Alison Freitas Silva (UNIPAMPA)**
 - <http://lattes.cnpq.br/7511013446343990>
 - **Dr. Jean Felipe Cheiran (UNIPAMPA)**
 - <http://lattes.cnpq.br/9076894715365563>

Direções futuras: Interface

- Apoio ao **aprendizado**:
 - Aperfeiçoar o **layout**
 - Revisar e ampliar as **instruções**
 - Adicionar **métricas** (e.g. consumo de recursos)
 - Integração com **Mlflow** (e.g., gráficos dinâmicos e personalizáveis)
- Revisar e melhorar a **usabilidade** geral
- Ajustes de **desempenho**
- **Compatibilidade** com múltiplos tipos de telas
- **Integração** com serviços de *chatbot*

Direções futuras: Cloud AutoDroid



- **Auto-escalabilidade** utilizando computação em Nuvem
- **Sistema de notificação** (e-mail, SMS, WhatsApp, push, etc.)
- **Dashboard administrativo** com estatísticas em tempo real
- **Monitoramento** (e.g., **Mlflow**) dos trabalhos sendo executados
- Estimativas/**cálculos dos custos** de processamento
- **Documentação** via OpenAPI



Direções futuras: MalSynGen



- **Integração** com serviços de *chatbot*
- Generalização da ferramenta **para outros conjuntos de dados**
- **Resumo** dos logs de saída
- Inclusão de mais **métricas de avaliação**
- **Comparação** com outras ferramentas
- **Generalização da ferramenta** para outros conjuntos de dados

Demo Day Malware DataLab

- Papers
- Slides (apresentações)
- Histórico das evoluções
- Vídeos técnicos



<https://drive.google.com/drive/folders/1gHqODJZ7AIOGII6hWm290GbmrGQTAZ0t?usp=sharing>



Obrigado!



GT Malware DataLab

<https://mdl.unihacker.club>

Demo Day

29 de novembro de 2024



Video

(overview da plataforma)

