DETEÇÃO DE BOTS NAS REDES SOCIAIS: MÉTODOS, DESAFIOS E EVOLUÇÃO

Amostragem e Fontes de Informação

Grupo 12

iscte INSTITUTO UNIVERSITÁRIO DE LISBOA

2022/2023

INTRODUÇÃO

INTRODUÇÃO

01
02
03
04
05
06
07

#1 Redes sociais desempenham um papel fundamental na disponibilização de informação na era digital.

#2 Aumento no número de usuários também pode atrair a atenção de indivíduos mal-intencionados.

INTRODUÇÃO

01
02
03
04
05
06
07

#3 Com o aumento da popularidade destas plataformas, as atenções começam a centrar-se em torno da cibersegurança.

#4 Os *bots* são a forma predominante de ***malware* nas redes sociais**.

INTRODUÇÃO

01
02
03
04
05
06
07

#5 Embora possam ser benignos, grande parte dos bots são maliciosos e são utilizados para atividades como **criação de contas falsas, *phishing*, *spam*** e para a **manipulação da opinião pública**.

#6 Os últimos anos têm sido marcados por avanços na detecção de *malware* em redes sociais, utilizando técnicas de aprendizagem supervisionada e não supervisionada.

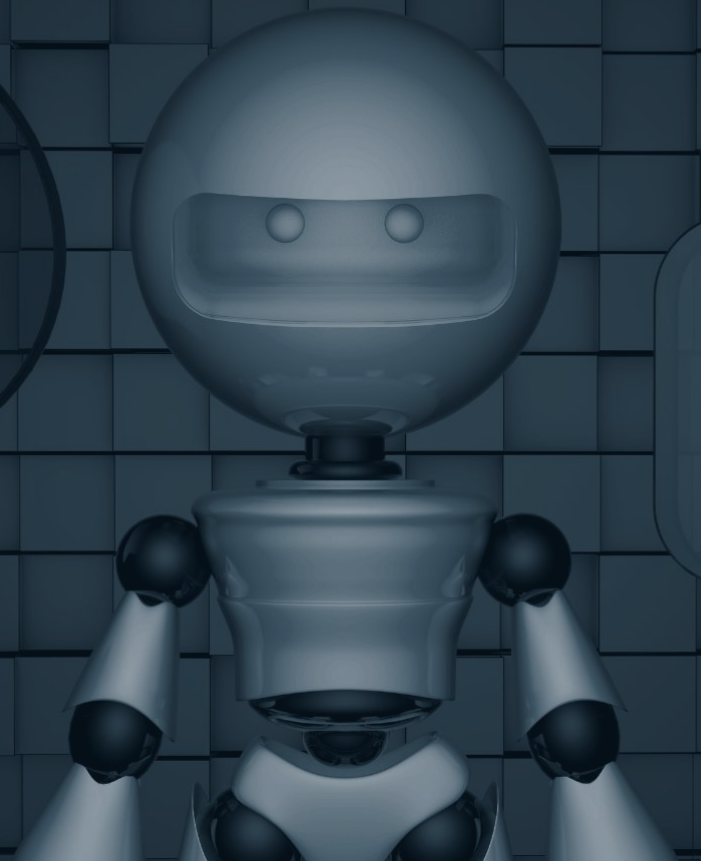
TIPOS DE *BOTS* NAS REDES SOCIAIS

TIPOS DE *BOTS* NAS REDES SOCIAIS

01
02
03
04
05
06
07

SOCIAL BOTS

Procuram reproduzir comportamentos humanos nas redes sociais.



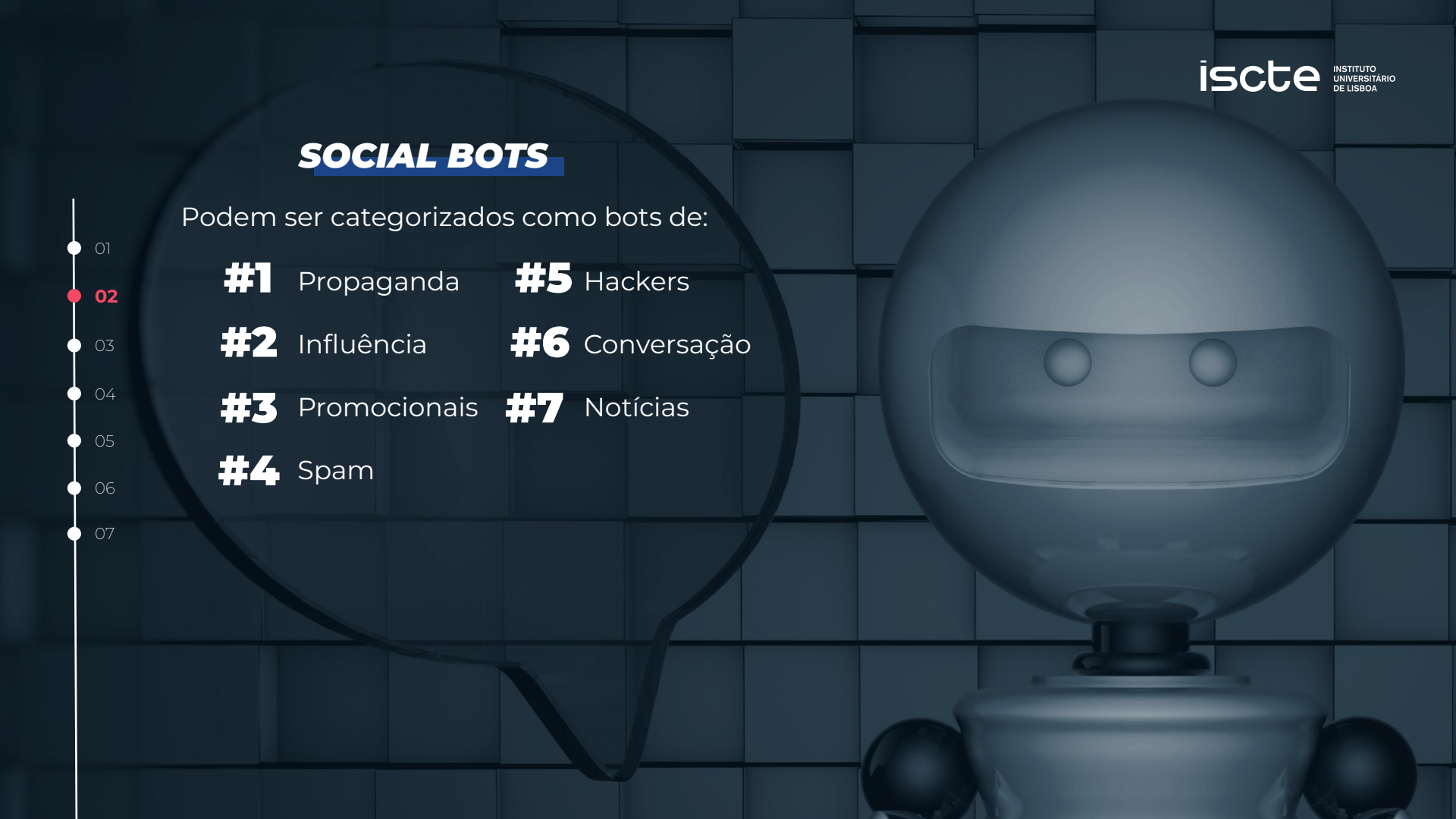
BOTNETS

Permitem a expansão dos ataques dos *social bots* para campanhas de spam, espionagem, phishing ou ransomware a larga escala

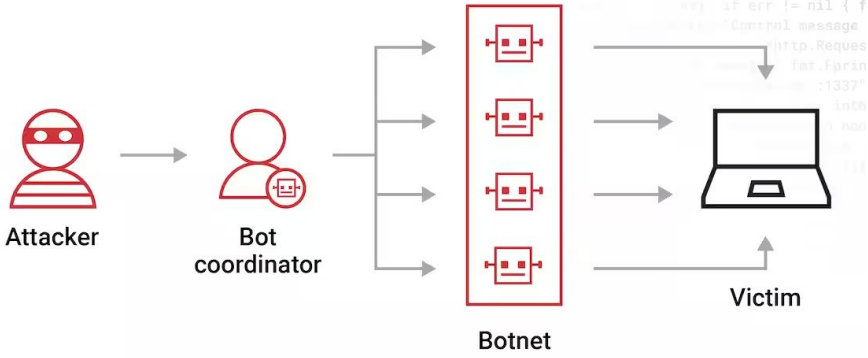
SOCIAL BOTS

Podem ser categorizados como bots de:

- **#1** Propaganda
- **#2** Influência
- **#3** Promocionais
- **#4** Spam
- **#5** Hackers
- **#6** Conversação
- **#7** Notícias



COMO FUNCIONA UMA *BOTNET*?



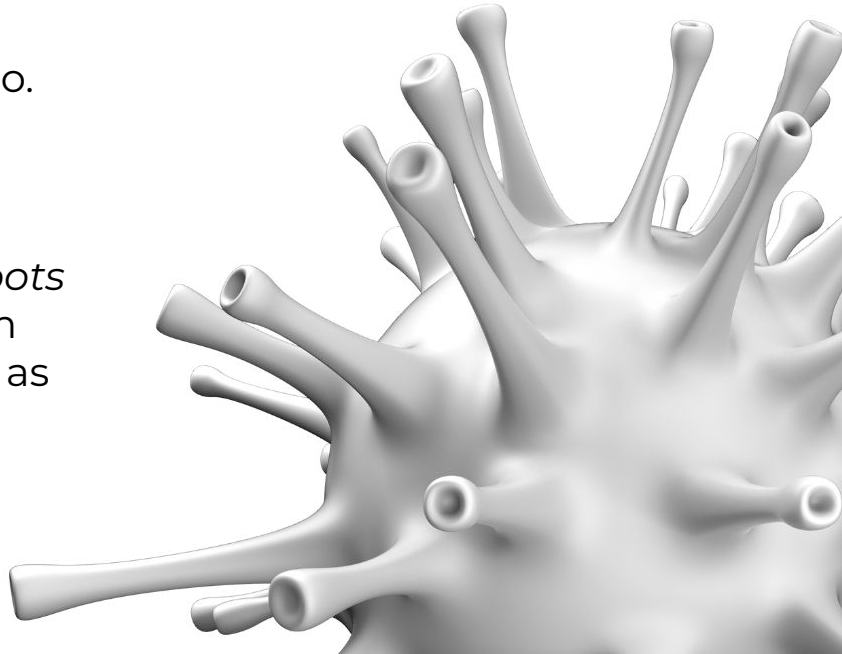
PRINCIPAIS IMPLICAÇÕES

O CASO DA COVID-19

01
02
03
04
05
06
07

#1 A COVID-19 ficou conhecida pelo termo "**infodemia**" pelo aumento de *social bots* que participavam ativamente no debate sobre as medidas de controlo pandémico.

#2 Assistiu-se a uma onda de desinformação, para a qual os *bots* contribuíram especialmente em temas mais controversos como as vacinas.



O CASO DA GUERRA ENTRE A RÚSSIA E A UCRÂNIA

Os bots foram um elemento-chave para **gerar awareness** para o problema da guerra entre a Rússia e a Ucrânia.

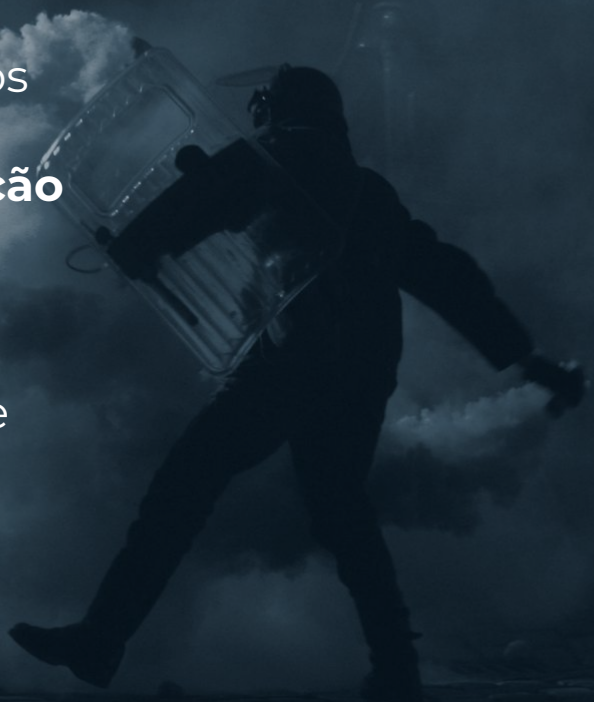
#1 A percentagem de publicações no Twitter de bots a favor do lado ucraniano foi quase **nove vezes maior** do que a do lado russo.

O CASO DA GUERRA ENTRE A RÚSSIA E A UCRÂNIA

01
02
03
04
05
06
07

#2 Os *social bots* amplificaram a voz dos ucranianos através da **condenação dos atos da Rússia** e da **manifestação de apoio à Ucrânia**.

#3 Do lado russo, os tweets de *bots* que publicavam com mais frequência focavam-se essencialmente em **conteúdo controverso** como a expansão da NATO para leste.



iscte INSTITUTO UNIVERSITÁRIO DE LISBOA

TÉCNICAS PARA A DETEÇÃO DE *BOTS*

TÉCNICAS PARA A DETEÇÃO DE BOTS

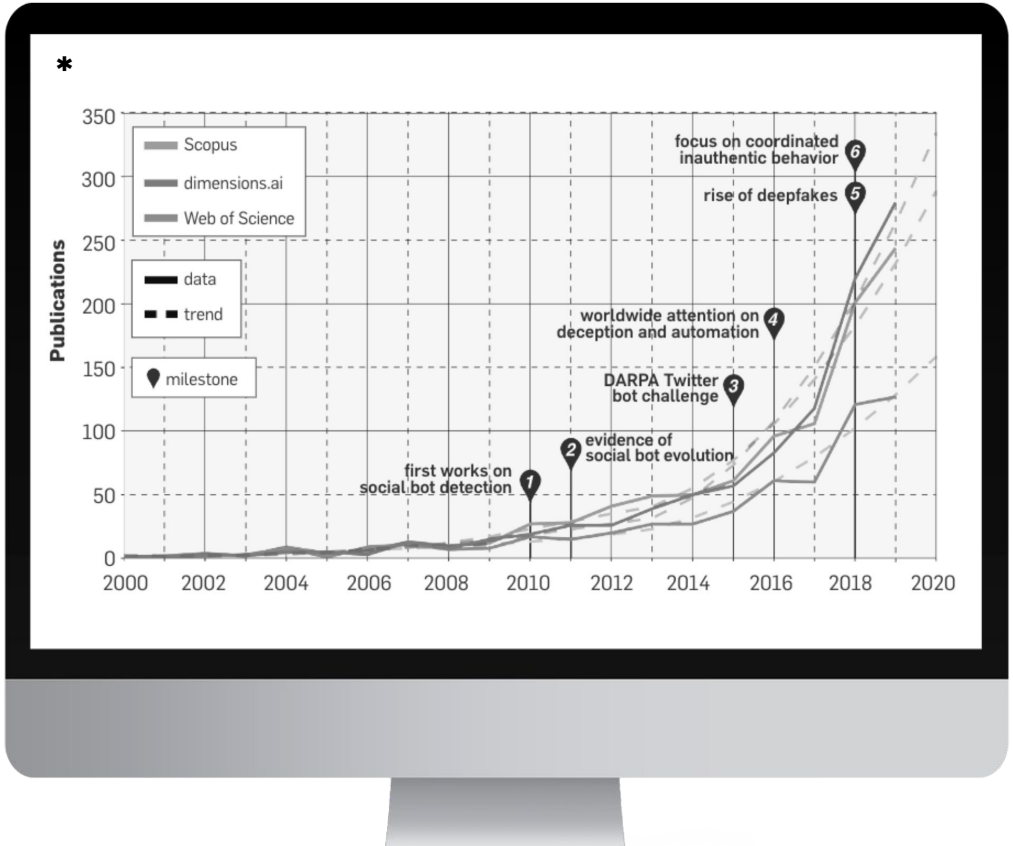
Machine Learning

Aprendizagem Supervisionada

Aprendizagem Não Supervisionada

TÉCNICAS DE APRENDIZAGEM SUPERVISIONADA

O primeiro trabalho desenvolvido especificamente para a deteção de contas automáticas em redes sociais surgiu em janeiro de 2010.



* Figura 1 - Publicações por ano sobre a caracterização, deteção e impacto estimado de social bots

METODOLOGIA DAS TÉCNICAS DE APRENDIZAGEM SUPERVISIONADA

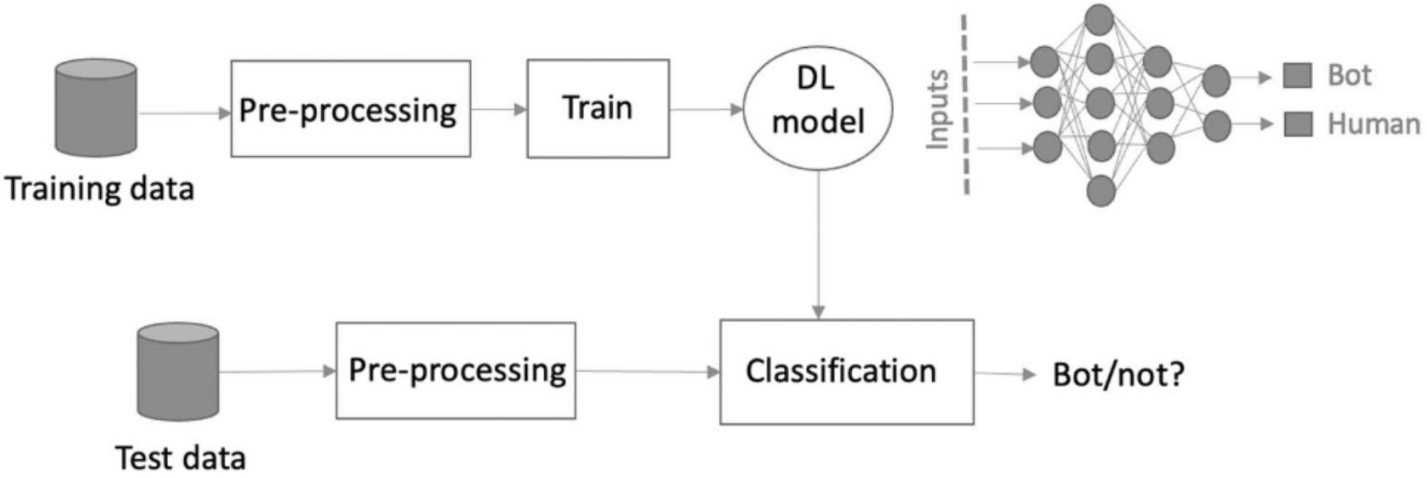
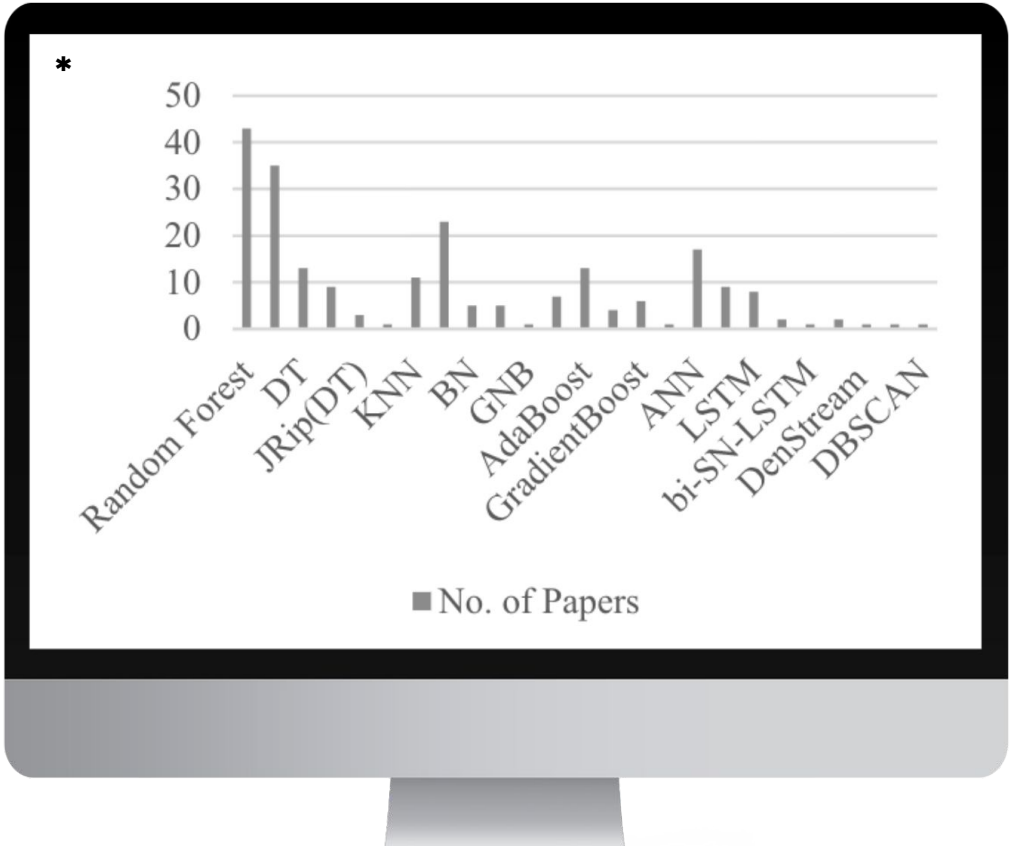


Figura 2 - Workflow de um modelo de classificação de um detetor de *bots*

TÉCNICAS DE APRENDIZAGEM SUPERVISIONADA

01
02
03
04
05
06
07

Os algoritmos utilizam características como a taxa de publicação de URLs em publicações, as taxas de interação, a dimensão do nome dos utilizadores, o rácio de seguidores-amigos e algumas análises de sentimento para identificar *bots* nas redes sociais.

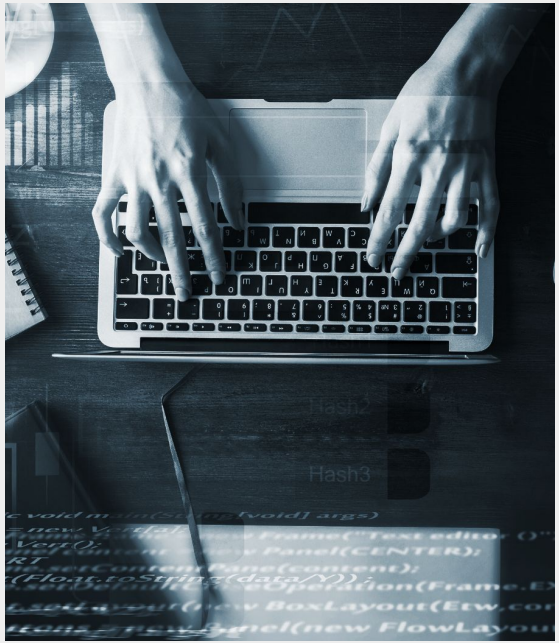


* Figura 3 - Principais algoritmos aplicados em papers sobre a deteção de bots em redes sociais

DESAFIOS À APLICAÇÃO DE TÉCNICAS SUPERVISIONADAS

01
02
03
04
05
06
07

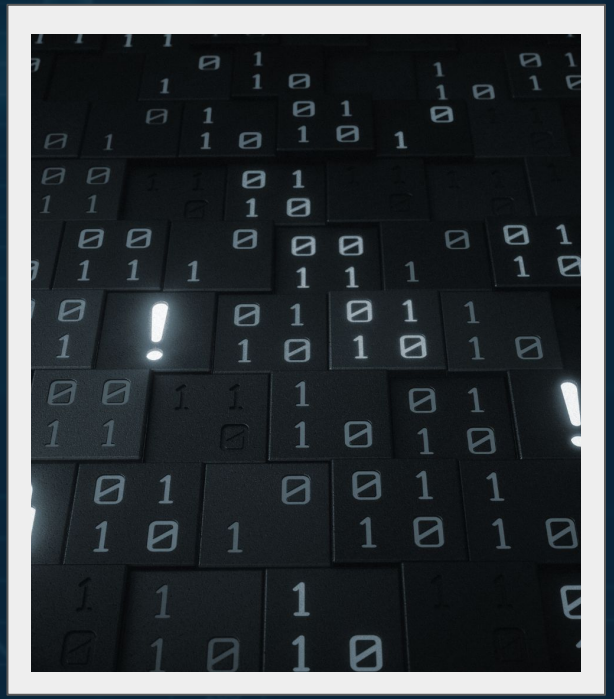
#1 Dataset utilizado para a validação dos dados previstos pelo modelo, que depende, em grande parte dos casos, de rótulos (bot/não bot) atribuídos por operadores humanos que procedem a uma análise manual, limitada e enviesada.



DESAFIOS À APLICAÇÃO DE TÉCNICAS SUPERVISIONADAS

01
02
03
04
05
06
07

#2 A produção de contramedidas por parte de quem desenvolve social bots tem vindo a dificultar esta tarefa de classificação.



DESAFIOS À APLICAÇÃO DE TÉCNICAS SUPERVISIONADAS

#3 Atualmente, os bots dispõem de fotografias de perfil, nomes e informações roubadas, vários seguidores reais e onde as mensagens maliciosas passam despercebidas entre inúmeras neutras, que são espalhadas através das botnets, o que dificulta a tarefa de os distinguir de utilizadores legítimos.

TÉCNICAS DE APRENDIZAGEM NÃO SUPERVISIONADA

01
02
03
04
05
06
07

Os métodos de aprendizagem automática não supervisionados que utilizam os dados da cronologia de uma determinada rede social podem oferecer uma performance comparável ou até melhor que os métodos supervisionados, **com menos enviesamento** por parte de operadores humanos e com um **custo de complexidade mais reduzido**.

TÉCNICAS DE APRENDIZAGEM NÃO SUPERVISIONADA

01
02
03
04
05
06
07

Estas técnicas consistem na criação de **clusters entre contas legítimas e de bots**, através dos algoritmos como o K-means ou o Agglomerative clustering.

TÉCNICAS DE APRENDIZAGEM SUPERVISIONADA VS NÃO SUPERVISIONADA

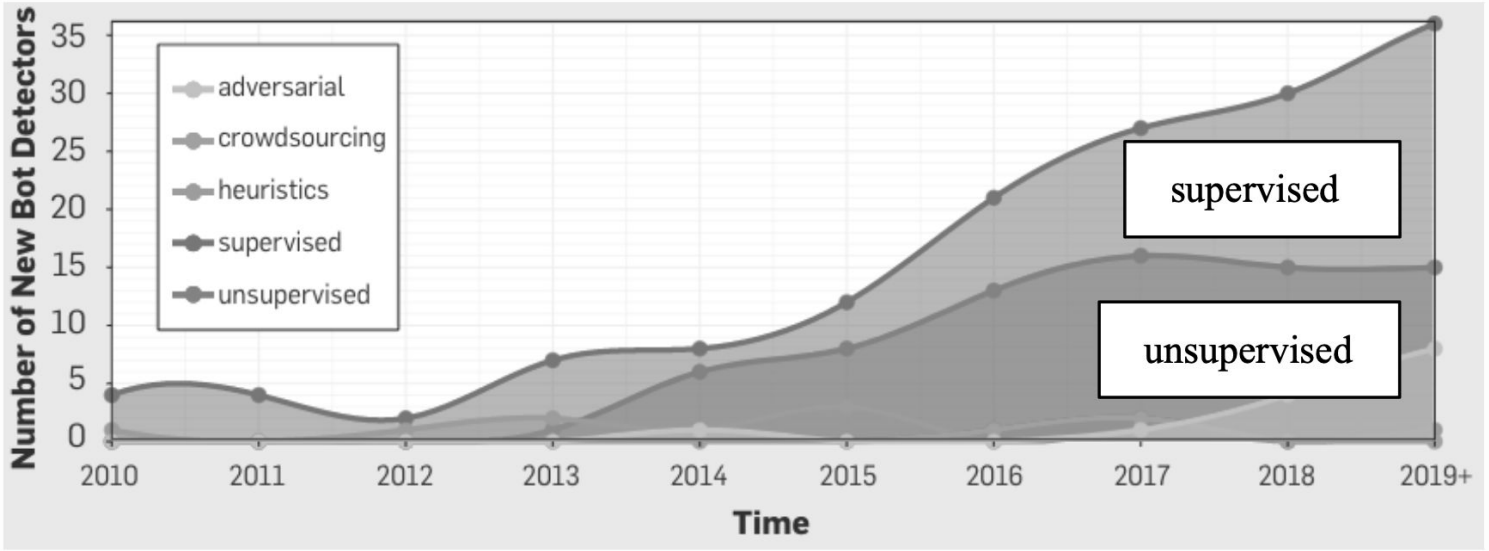


Figura 4 - Número de novos detetores de bots

INCORPORAÇÃO DE MECANISMOS DE DETEÇÃO NA ESCUTA ATIVA

01
02
03
04
05
06
07

De acordo com uma entrevista realizada a João Santos Silva, consultor de Assuntos Públicos da empresa LLYC, a escuta ativa nas redes sociais é essencial para identificar os *pain points* sobre o qual surge o conflito e os principais *opinion makers* que se manifestam sobre o tema.

INCORPORAÇÃO DE MECANISMOS DE DETEÇÃO NA ESCUTA ATIVA

01
02
03
04
05
06
07

Dado o uso pernicioso da IA na produção de conteúdos massificados para fins de manipulação, como se viu nas eleições norte-americanas de 2016, onde um quinto de todos os tweets sobre as eleições presidenciais foram publicados por bots, será necessário que as plataformas de escuta ativa sinalizem rapidamente quais os conteúdos produzidos por bots.

iscte INSTITUTO UNIVERSITÁRIO DE LISBOA

INCORPORAÇÃO DE MECANISMOS DE DETEÇÃO NA ESCUTA ATIVA

01
02
03
04
05
06
07

A LLYC utiliza o Brandwatch para para se inteirar das conversas que decorrem no espaço social do Twitter, mapear comunidades e identificar potenciais Key Opinion Leaders (KOLs). A plataforma disponibiliza filtros de spam nas configurações de cada projeto e ferramentas de “Bot Detection”.

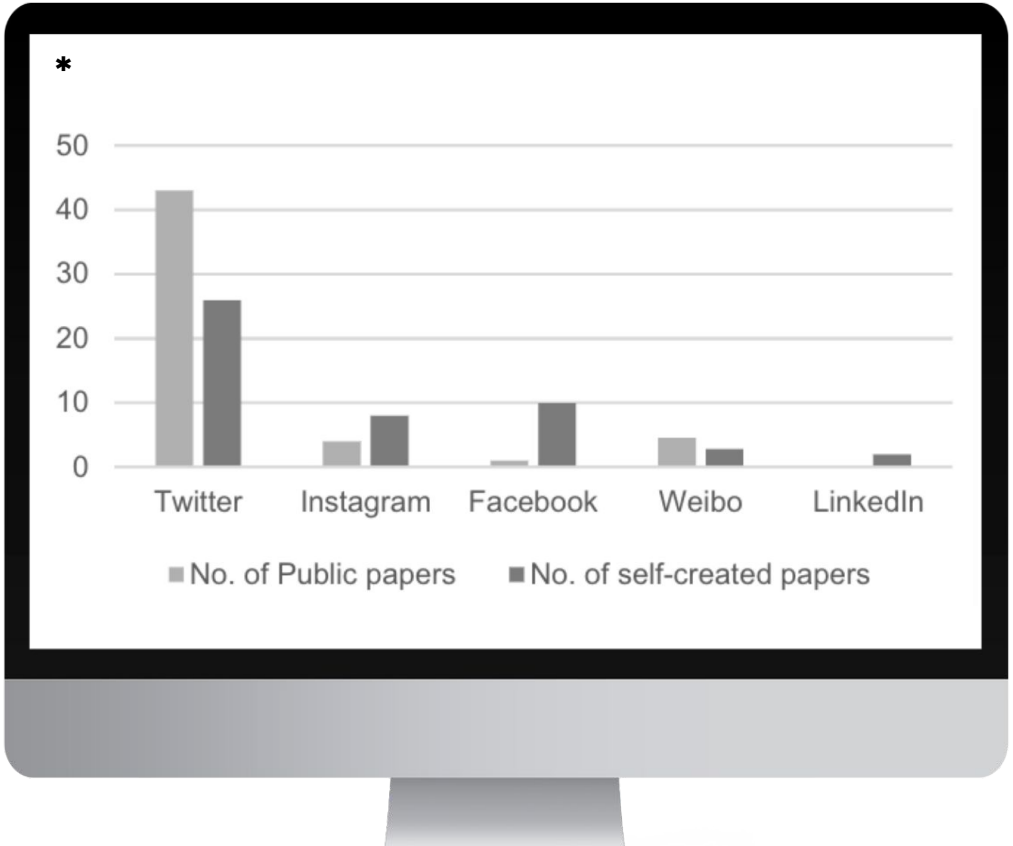
DESAFIOS E OPORTUNIDADES

DESAFIOS E OPORTUNIDADES

01
02
03
04
05
06
07

#1

A maior parte da investigação conduzida nesta área é feita na rede social Twitter.

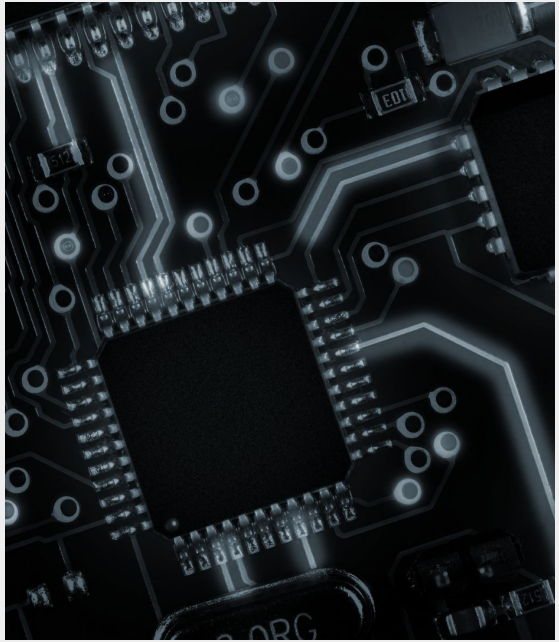


* Figura 5 - Distribuição de *datasets* por plataforma

DESAFIOS E OPORTUNIDADES

#2 A generalidade dos mecanismos de deteção de bots atuais segue um esquema reativo.

#3 A investigação neste setor está assente em pressupostos estáticos e neutros.



PRINCIPAIS CONCLUSÕES

PRINCIPAIS CONCLUSÕES

01
02
03
04
05
06
07

#1

Categorização de bots: Social bots e botnets tiveram impacto na manipulação da opinião pública.

#2

Deteção de bots: Métodos de aprendizagem automática supervisionada são predominantes, mas apresentam algumas limitações.

PRINCIPAIS CONCLUSÕES

#3

Desafios e oportunidades: Expandir a pesquisa para analisar outras plataformas além do Twitter em Portugal e adotar uma atitude proativa na sinalização e deteção de *bots*.

#4

Importância contínua: O estudo dos bots em redes sociais é crucial para combater a desinformação e preservar a integridade das plataformas.

OBRIGADO

Amostragem e Fontes de Informação

Grupo 12