



# CRACKING THE CRAPTCHA

JUST BECAUSE

The slide features a dark grey background with several translucent, realistic-looking bubbles of various sizes scattered across the top and bottom edges. The bubbles have highlights and shadows, giving them a three-dimensional appearance.

# AGENDA

- INTRODUÇÃO
- QUANDO UM CAPTCHA É QUEBRÁVEL?
- UMA BREVE VISITA NO PROCESSO
- AUTOMATIZAÇÃO

# INTRODUÇÃO

O QUE É CAPTCHA?

1. **C**OMPLETE **A**UTOMATED **P**UBLIC **T**URING TEST TO TELL **C**OMPUTERS AND **H**UMANS **A**PART
2. TESTE PARA DIFERENCIAR UMA PESSOA DE UM ROBÔ
3. USADO PARA COMBATER SPAMS E AUTOMAÇÃO DE EXTRAÇÃO DE DADOS

# INTRODUÇÃO

A ESTÓRIA DE CARLOS, SAULO E HUMBERTO

1. CARLOS É O HERÓI DA GURIZADA
2. SAULO É O AMIGO POPULAR DE CARLOS
3. HUMBERTO É O CACHORRO DE CARLOS

A TRETA COMEÇA A PARTIR DO CONCURÇÃO, CONCURSO PROMOVIDO POR UMA RÁDIO LOCAL, ONDE A FOTO DO CÃO COM MAIOR NOTA MÉDIA, É A FOTO VENCEDORA.

AS FOTOS SÃO POSTADAS NO SITE DA RÁDIO, E PARA VOTAR, VOCÊ PRECISA SE CADASTRAR.

# INTRODUÇÃO

- ALGUNS TERMOS UTILIZADOS NA APRESENTAÇÃO:

1. SESSÃO – TODAS AS INTERAÇÕES ENTRE O USUÁRIO E O SERVIDOR: REQUISIÇÕES HTTP(S), COOKIES, ETC.
2. SOLUÇÃO – RESPOSTA CORRETA PARA DESAFIO DO SISTEMA DE CAPTCHA, DENTRO DE UMA SESSÃO
3. IMAGEM – FIGURA GERADA PELO SISTEMA DE CAPTCHA REPRESENTANDO UM CONJUNTO DE CARACTERES ALFANUMÉRICOS PARA DESCREVER A SOLUÇÃO

# INTRODUÇÃO

TRANSFORMAR DE UMA IMAGEM PARA UM TEXTO É UM PROCESSADO CHAMADO **OCR** (OPTICAL CHARACTER RECOGNITION).

COMO FAZER OCR DESSA IMAGEM?

1. APRENDER COMPUTER VISION (?!?)
2. UTILIZAR ALGUMA FERRAMENTA JÁ PRONTA (TESSERACT)
3. PROCURAR UM TUTORIAL DE QUEBRA DE CAPTCHA ONLINE

Information



# QUANDO UM CAPTCHA É QUEBRÁVEL?

- AO FAZER RELOAD NA IMAGEM (F5):
  1. ELA PERMANECE A MESMA?
  2. MUDA COMPLETAMENTE?
  3. VARIA A POSIÇÃO DAS LETRAS?
- EXISTE UM TEMPO PARA QUE A URL DA IMAGEM PERMANEÇA VÁLIDA?
  1. EXISTÊNCIA DE UMA BASE DE DADOS PARA ASSOCIAÇÃO DE SESSÃO/SOLUÇÃO
  2. TEMPO MÁXIMO PARA CAPTURA DE IMAGENS
  3. EXISTÊNCIA DE PRAZO DE VALIDADE

# QUANDO UM CAPTCHA É QUEBRÁVEL?

- APÓS UMA TENTATIVA (ERRADA), O CAPTCHA:
  1. PERMANECE O MESMO?
  2. MUDA COMPLETAMENTE?
- É POSSÍVEL IDENTIFICAR SE O CAPTCHA É ASSOCIADO A UMA SESSÃO?
  1. ANÁLISE DE COOKIES ASSOCIADOS
  2. É POSSÍVEL INJETAR A ÚLTIMA SESSÃO PARA MANTER A MESMA SOLUÇÃO?
- EXISTE UM NÚMERO LIMITE DE TENTATIVAS?
  1. RESTRINGE A MARGEM DE ERRO
  2. UM MAIOR NÚMERO DE IMAGENS DEVE SER CAPTURADAS PARA PROCESSAMENTO





AVISO: A REVISTA H2HC TEM UM ARTIGO INTEIRO SOBRE O PROCESSO DE DECODIFICAÇÃO, COM CÓDIGO FONTE, E AS IMAGENS EXTRAÍDAS PARA AMOSTRAGEM.

# DECODIFICANDO

MELHOR MANEIRA DE COMEÇAR: **UMA AMOSTRAGEM**

50 SESSÕES, 50 IMAGENS CADA SESSÃO

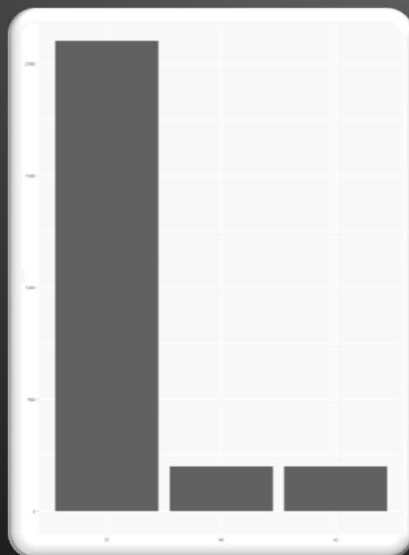
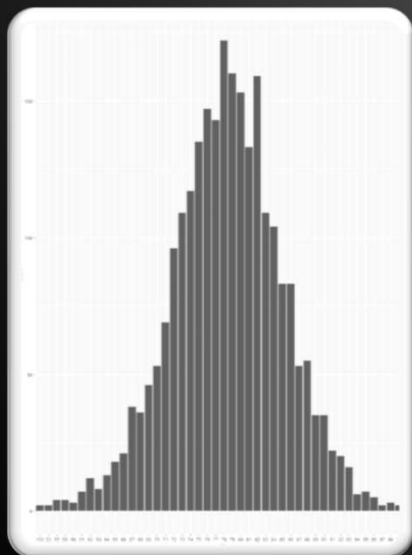


# DECODIFICANDO

1. NORMALIZA-SE A IMAGEM (RESOLUÇÃO)
2. PRÉ-FILTRO (EU CHAMO DE MASCARAMENTO)
3. SEPARAÇÃO DAS LETRAS
4. ENVIO DE CADA LETRA PARA O TESSERACT



# DECODIFICANDO



HISTOGRAMA DA AMOSTRAGEM (2500 IMAGENS) NA LARGURA E ALTURA.

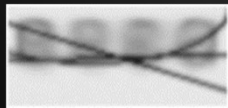
NORMALIZAÇÃO DA IMAGEM:

- MANTER TODAS AS IMAGENS NA MESMA RESOLUÇÃO (LARGURA E ALTURA)

# DECODIFICANDO

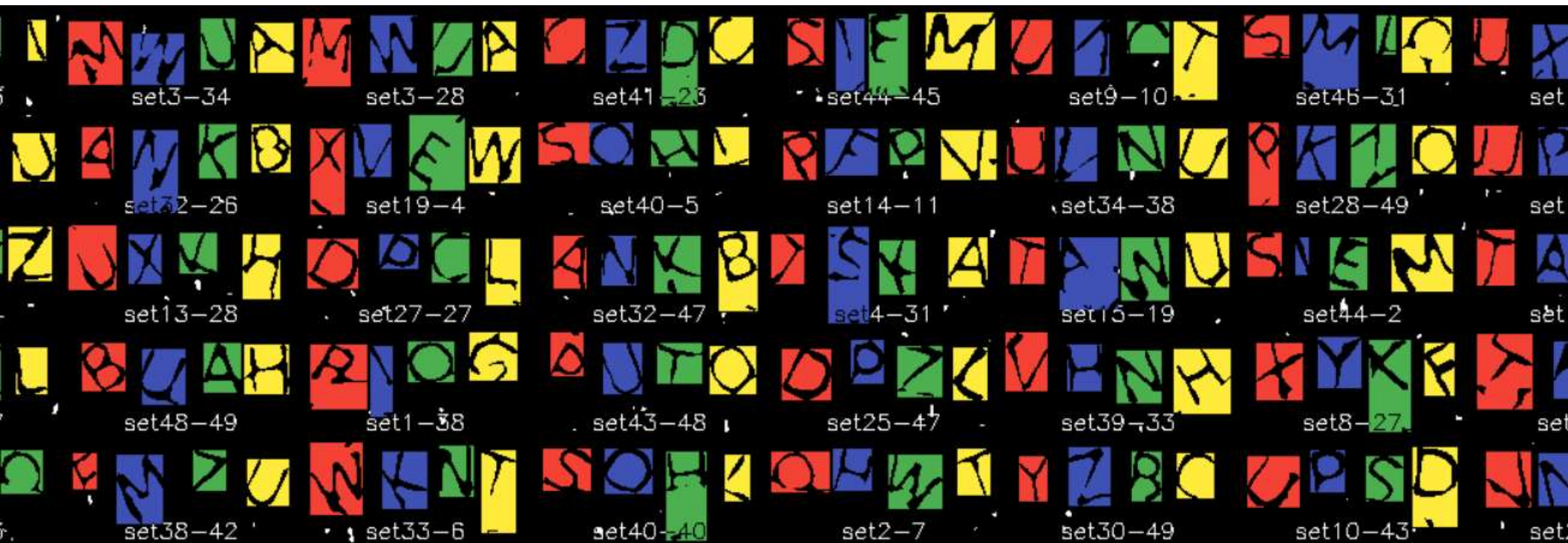
## PRÉ-FILTRO:

- CONSIDERAR UMA AMOSTRAGEM E RETIRAR ARTEFATOS DA IMAGEM
- EXEMPLO DA MÉDIA DE TODAS AS IMAGENS DA AMOSTRAGEM SUBTRAÍDAS DAS IMAGENS



MÉDIA





# SEPARAÇÃO DAS LETRAS: O MAIS COMPLICADO

## DECODIFICANDO

		QUANTAS LETRAS QUE O SOFTWARE CONSEGUIU RECONHECER																										
IMAGENS	SOLUCAO	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	DECODED
0	R			2	4					1	2	15				1	1	1	6									15
1	J	5								1	2	15				2					1	1						27
2	H	2								18					7		1	1								2		31
3	W									1				1	9			2		2		1	15			1	32	
4	S		15			3	1									3	1	4		1		1				1	30	
5	J	1								2	7	1				2		1					1	1			11	
6	J	2									13		1		2		1					1	1				21	
7	M	1								2						24	1		2				2	1	3		36	
8	Y										4	1											21			4	30	
9	A	26																									26	
10	P				11									5	2	2									1		21	
11	X					1						1										1		5	7		15	
12	H	2								16				1								2	2				23	
13	X									1		1		1								6		4	11		23	
14	F			10			8															1			3		22	
15	A	19			3		1								1		4										28	
16	O													13													13	
17	I	2								3									1				1				7	
18	C	1		27			6									2											36	
19	V					1					1		1	1				1				1		10	7	1	24	
20	F			13		1	5									1								3		2	25	
21	C			20			9					1											1				31	
22	Y										2	1	1										16			3	23	
23	C	1		32			5									1											41	
24	Y									4		2											20			4	30	
25	P	2			12									13		1						1					29	
26	L				1					1		13	1										1	1	1	2	23	
27	P				20		1							5	1	1						1					29	
28	K										30	1											1				32	
29	N	1											6	13							2				3		25	
30	Z																									20	20	
31	A	18			2	1										7	1										29	
32	N													4	9		5							1	3		22	

		QUANTAS LETRAS QUE O SOFTWARE CONSEGUIU RECONHECER																										
IMAGENS	SOLUCAO	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	DECODED
0	T																										15	
1	B		2																								27	
2	Q			1																							31	
3	M	1												25	1												32	
4	T																										30	
5	E																										11	
6	R																										21	
7	X																										36	
8	X																										30	
9	U																										26	
10	U																										21	
11	C	2		15																							15	
12	U																										23	
13	U																										23	
14	P																										22	
15	T																										28	
16	M																										13	
17	M																										7	
18	S																										36	
19	X																										24	
20	G																										25	
21	A																										31	
22	G																										23	
23	M																										41	
24	M																										30	
25	D																										29	
26	Z																										23	
27	D																										29	
28	P																										32	
29	V	1																									25	
30	Y																										20	
31	R																										29	
32	A																										22	

SEGUNDA LETRA

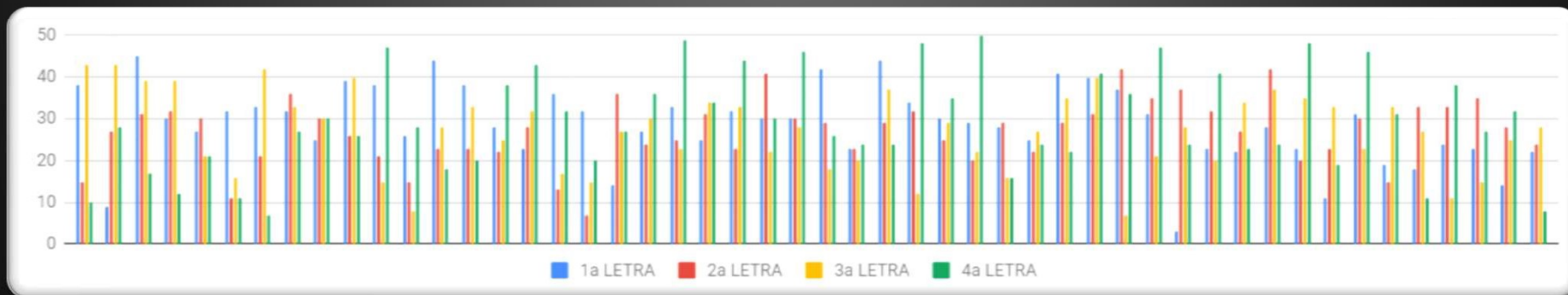
PRIMEIRA LETRA

# ENVIO DE CADA LETRA PARA O TESSERACT

DECODIFICANDO

# DECODIFICANDO

- DECODIFICAÇÃO DA IMAGEM É PROCESSO ANALÓGICO:
  - FUNÇÃO DECODIFICA(IMAGEM) RETORNA QUATRO LETRAS, PODENDO SER VAZIO (NULO) QUANDO NÃO É POSSÍVEL DECODIFICAR
  - ALTO ÍNDICE DE ERRO
- É UM PROCESSO ESTATÍSTICO:
  - SE O NÚMERO DE IMAGENS É INFINITO, É CERTO QUE VOCÊ ACERTARÁ O CAPTCHA
  - SE O NÚMERO DE TENTATIVAS É INFINITO, VOCÊ TAMBÉM PODE USAR FORÇA BRUTA 😊



DECODIFICANDO



ANÁLISE DE CONTEÚDO DE PALAVRAS EM CONJUNTOS DE DADOS																												
IMAGEM	PRIMEIRA									SEGUNDA						TERCEIRA					QUARTA							
	TESSERACT			None	T	I	Y	S	K		None	R	O	Q	E	D	B		None	C	G	F	Q		None	K	I	V
				0	0	0	0	0	0		0	0	0	0	0	0		0	0	0	0	0		0	0	0	0	
set0-0	T	R	C	0	1	0	0	0	0		0	1	0	0	0	0		0	1	0	0	0		1	0	0	0	
set0-1	T	R	G	0	2	0	0	0	0		0	2	0	0	0	0		0	1	1	0	0		2	0	0	0	
set0-10	T		C	K	0	3	0	0	0		1	2	0	0	0	0		0	2	1	0	0		2	1	0	0	
set0-11			C		1	3	0	0	0		2	2	0	0	0	0		0	3	1	0	0		3	1	0	0	
set0-12	T	O	C		1	4	0	0	0		2	2	1	0	0	0		0	4	1	0	0		4	1	0	0	
set0-13	T	R	C		1	5	0	0	0		2	3	1	0	0	0		0	5	1	0	0		5	1	0	0	
set0-14	I		F		1	5	1	0	0		3	3	1	0	0	0		0	5	1	1	0		6	1	0	0	
set0-15		R			2	5	1	0	0		3	4	1	0	0	0		1	5	1	1	0		7	1	0	0	
set0-16	T	R	C		2	6	1	0	0		3	5	1	0	0	0		1	6	1	1	0		8	1	0	0	
set0-17	T	R	C		2	7	1	0	0		3	6	1	0	0	0		1	7	1	1	0		9	1	0	0	
set0-18	T		C		2	8	1	0	0		4	6	1	0	0	0		1	8	1	1	0		10	1	0	0	
set0-19	T	R	C		2	9	1	0	0		4	7	1	0	0	0		1	9	1	1	0		11	1	0	0	
set0-2	T	Q			2	10	1	0	0		4	7	1	1	0	0		2	9	1	1	0		12	1	0	0	
set0-20		R			3	10	1	0	0		4	8	1	1	0	0		3	9	1	1	0		13	1	0	0	
set0-21	T	R	C		3	11	1	0	0		4	9	1	1	0	0		3	10	1	1	0		14	1	0	0	
set0-22	T	R	C		3	12	1	0	0		4	10	1	1	0	0		3	11	1	1	0		15	1	0	0	
set0-23			C		4	12	1	0	0		5	10	1	1	0	0		3	12	1	1	0		16	1	0	0	
set0-24	T	R	C		4	13	1	0	0		5	11	1	1	0	0		3	13	1	1	0		17	1	0	0	
set0-25	Y		C	I	4	13	1	1	0		6	11	1	1	0	0		3	14	1	1	0		17	1	1	0	
set0-26			C	I	5	13	1	1	0		7	11	1	1	0	0		3	15	1	1	0		17	1	2	0	
set0-27		O		I	6	13	1	1	0		7	11	2	1	0	0		4	15	1	1	0		17	1	3	0	
set0-28			C		7	13	1	1	0		8	11	2	1	0	0		4	16	1	1	0		18	1	3	0	
set0-29	T	O	C		7	14	1	1	0		8	11	3	1	0	0		4	17	1	1	0		19	1	3	0	
set0-3	T	R	C		7	15	1	1	0		8	12	3	1	0	0		4	18	1	1	0		20	1	3	0	
set0-30	T		C		7	16	1	1	0		9	12	3	1	0	0		4	19	1	1	0		21	1	3	0	
set0-31	T	Q	C		7	17	1	1	0		9	12	3	2	0	0		4	20	1	1	0		22	1	3	0	
set0-32	T		C	I	7	18	1	1	0		10	12	3	2	0	0		4	21	1	1	0		22	1	4	0	
set0-33	T		C		7	19	1	1	0		11	12	3	2	0	0		4	22	1	1	0		23	1	4	0	
set0-34	S	E	C		7	19	1	1	1	0		11	12	3	2	1	0		4	23	1	1	0		24	1	4	0
set0-35			C		8	19	1	1	1	0		12	12	3	2	1	0		4	24	1	1	0		25	1	4	0
set0-36	T	R	Q		8	20	1	1	1	0		12	13	3	2	1	0		4	24	1	1	1		26	1	4	0

## DECODIFICANDO: RESULTADOS

	1ª LETRA	2ª LETRA	3ª LETRA	4ª LETRA
IMAGENS DECODIFICADAS	57%	54%	54%	59%
DECODIFICOU A SOLUCAO?	100%	98%	98%	100%
SOLUÇÃO É A MAIS FREQUENTE?	76%	70%	86%	88%

# DECODIFICANDO: RESUMO

DE POSSE DA AMOSTRA, PASSOS PARA DECODIFICAÇÃO:

1. NORMALIZAÇÃO
2. FILTRO DA IMAGEM
3. SEPARAÇÃO DAS LETRAS
4. ENVIO DE CADA LETRA PARA OCR

# DECODIFICANDO: QUIZ

- [HTTPS://WWW12.XXXX.LEG.BR/ECIDADANIA/@@CAPTCHA/IMAGE](https://www12.xxxx.leg.br/ecidadania/@@captcha/image)



# DECODIFICANDO: QUIZ

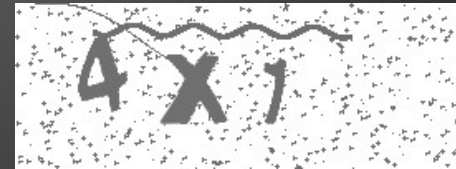
## MATH EQUATIONS




$7 \times 4$



$9 - 7$



$4 \times 1$



$7 + 1$



$8 + 4$



$9 + 2$

# DECODIFICANDO: QUIZ

A BIT DIFFERENT



# DECODIFICANDO: ALTERNATIVA

USO DE IMAGEM RESTRINGE PESSOAS COM DIFICULDADES DE VISÃO. QUEBRAR AUDIO RECITANDO AS LETRAS, TEORICAMENTE, É MAIS FACIL.

TEMA DE CASA 😊

# REFERENCIAS

- [HTTPS://MEDIUM.COM/@AGEITGEY/HOW-TO-BREAK-A-CAPTCHA-SYSTEM-IN-15-MINUTES-WITH-MACHINE-LEARNING-DBEBB035A710](https://medium.com/@ageitgey/how-to-break-a-captcha-system-in-15-minutes-with-machine-learning-dbebb035a710)



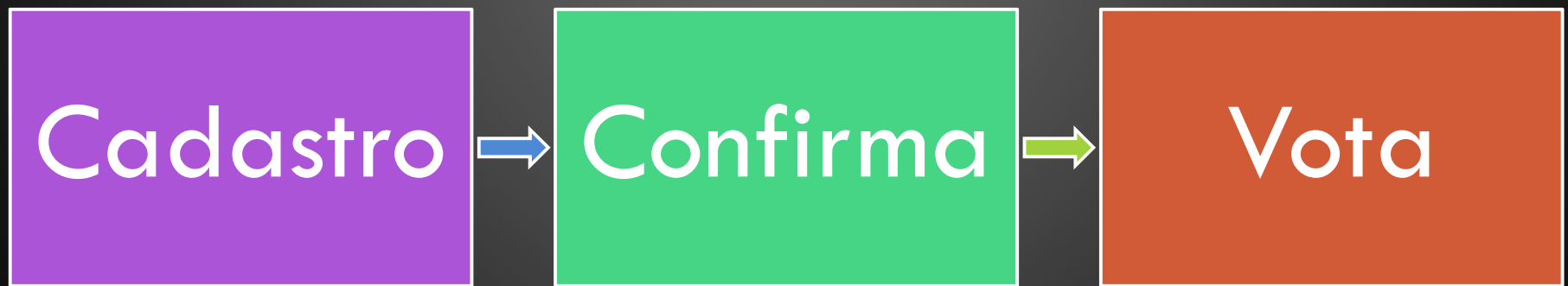
The background is a dark grey/black gradient with several translucent, realistic-looking bubbles of various sizes scattered across it. The bubbles have highlights and shadows, giving them a 3D effect. They are primarily located in the top-left and bottom-right corners, with a few smaller ones in the middle.

# E ERA ISSO

GUSTAVO SCOTTI

[CSH@OUTLOOK.COM](mailto:CSH@OUTLOOK.COM)

# UM SISTEMA DE VOTOS ONLINE



- NOME
- E-MAIL
- SENHA
- CAPTCHA

- RECEBE E-MAIL
- CLICA LINK
- TESTA LOGIN

- LOGIN
- POST DO VOTO

# PROTEGENDO CONTRA?

- IDÉIAS?