

¡BIENVENIDOS!

Iniciaremos en unos minutos...



IA/ML/DL

**Inteligencia
Artificial**

The diagram consists of three concentric circles. The outermost circle is light blue and contains the text 'Inteligencia Artificial'. Inside it is a medium blue circle containing 'Aprendizaje Automatizado'. The innermost circle is a darker blue and contains 'Aprendizaje Profundo'. To the right of these circles are three rounded rectangular boxes, each containing a description of the corresponding level of AI.

**Aprendizaje
Automatizado**

**Aprendizaje
Profundo**

Incorporar comportamiento
humano inteligente

Aprende y mejora automáticamente
de la experiencia*

Utiliza algoritmos complejos y redes
neuronales para entrenar un
modelo.

El impacto de la IA en el mundo

Today

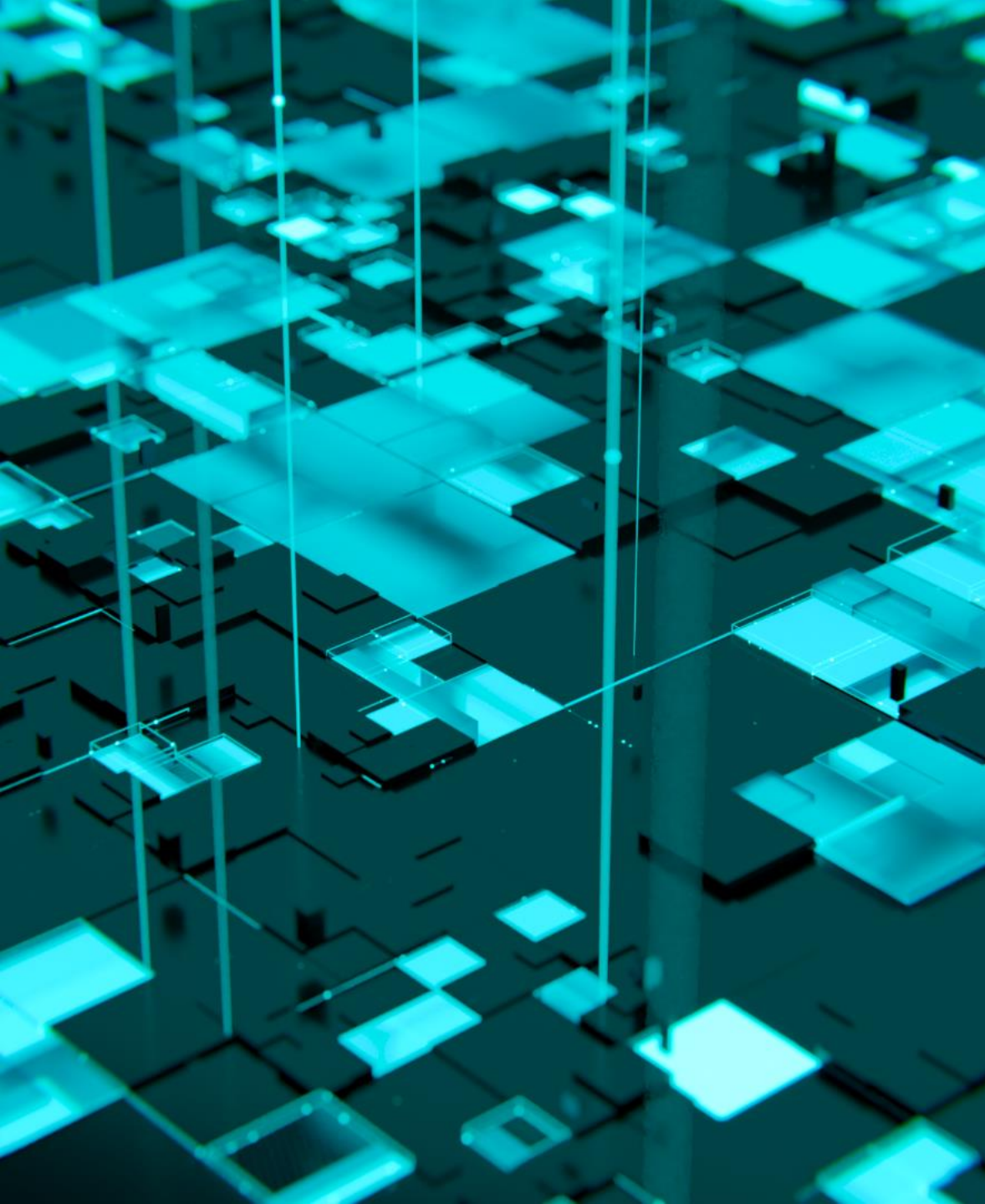
Hablar con los datos del negocio

Productividad sobrada

Constucción de estrategia

Mitigación de Interrupciones y
Riesgo operativo

Panorama de amenazas que usan IA para 2024



Podemos generalizar las tareas donde la
IA Ofensiva es más común en las
siguientes:

PREDICCIÓN

GENERACIÓN

ANÁLISIS

DEVOLUCIÓN

TOMA DE DECISIONES



Podemos generalizar las tareas donde se ataca a la IA con las siguientes:

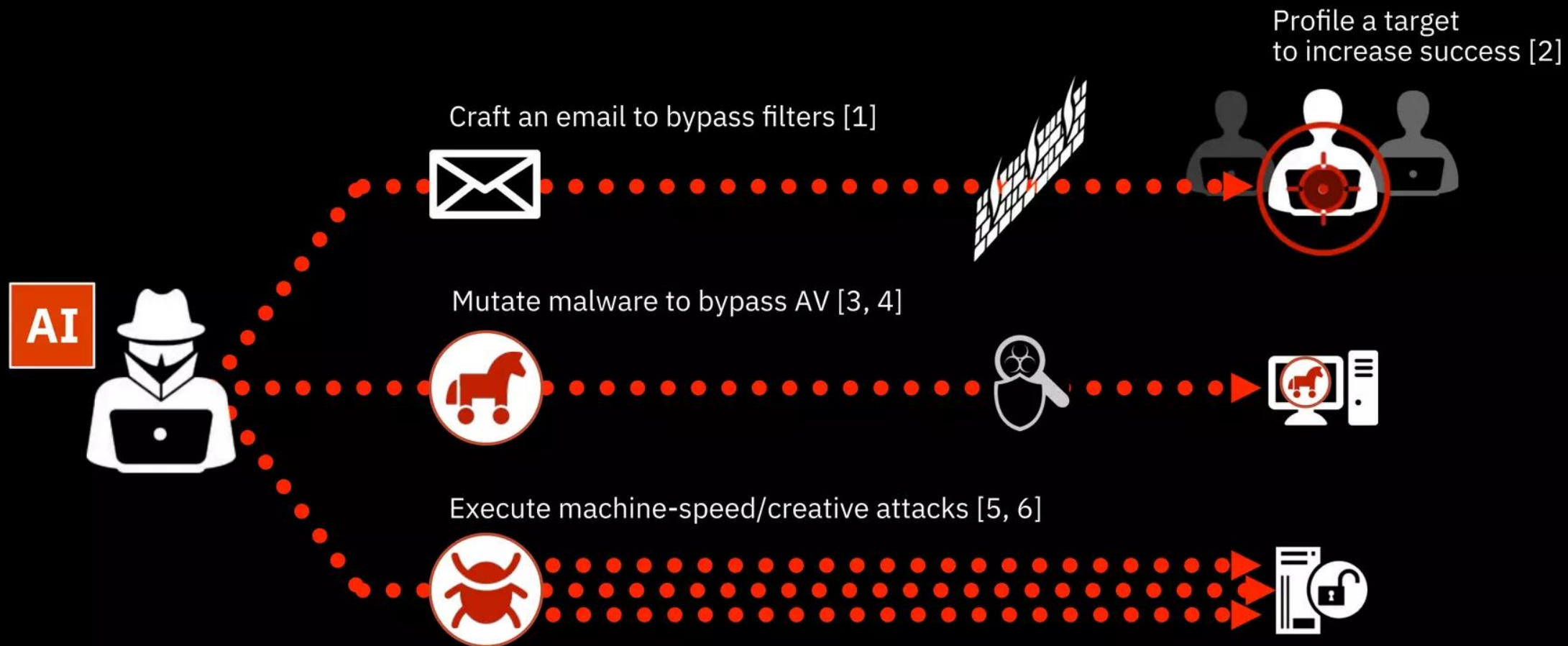
ANÁLISIS DE LAS RESPUESTAS

PARÁMETROS DEL MODELO

CÓDIGO DE ENTRENAMIENTO

MODIFICACION DATOS

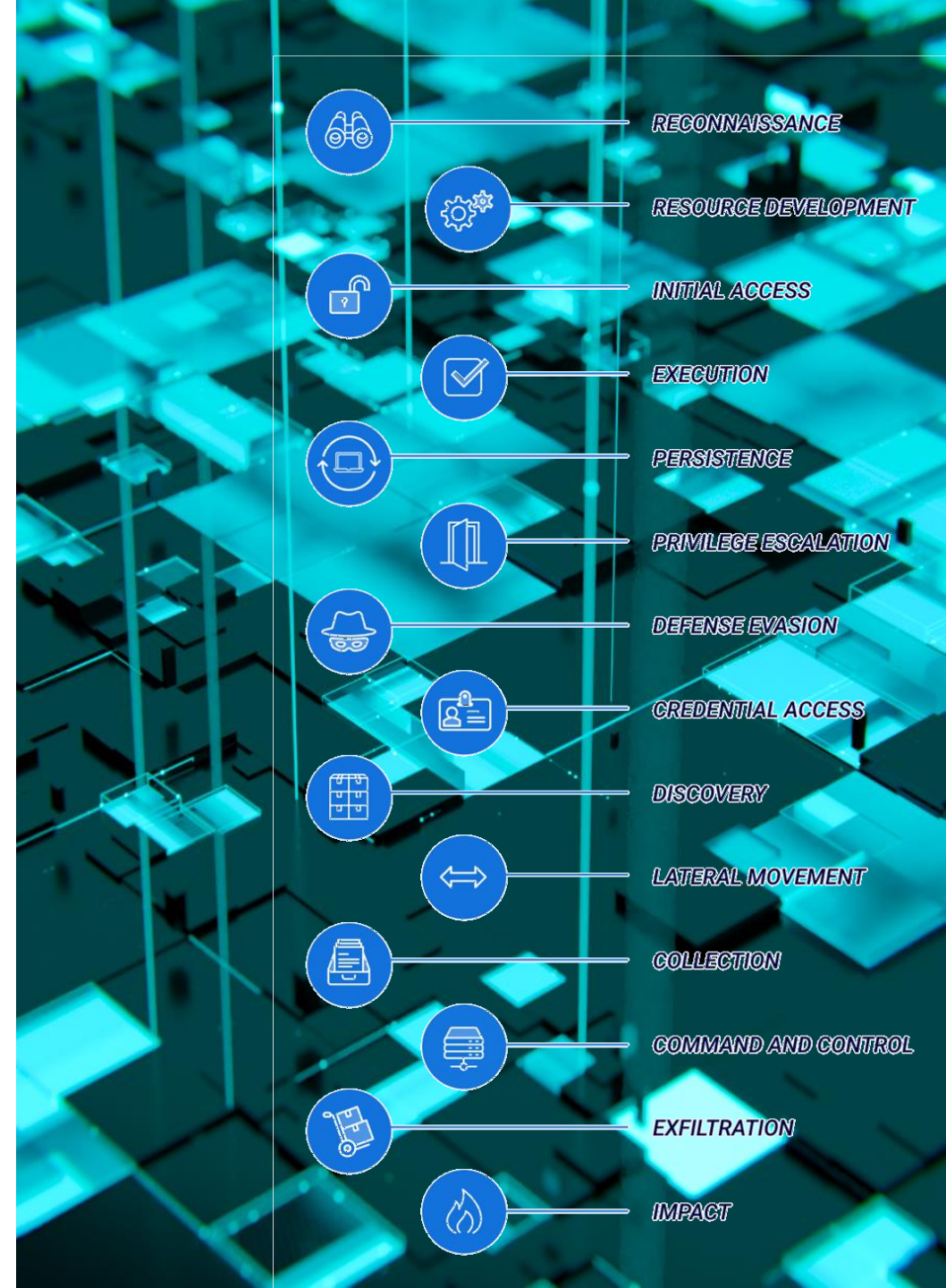
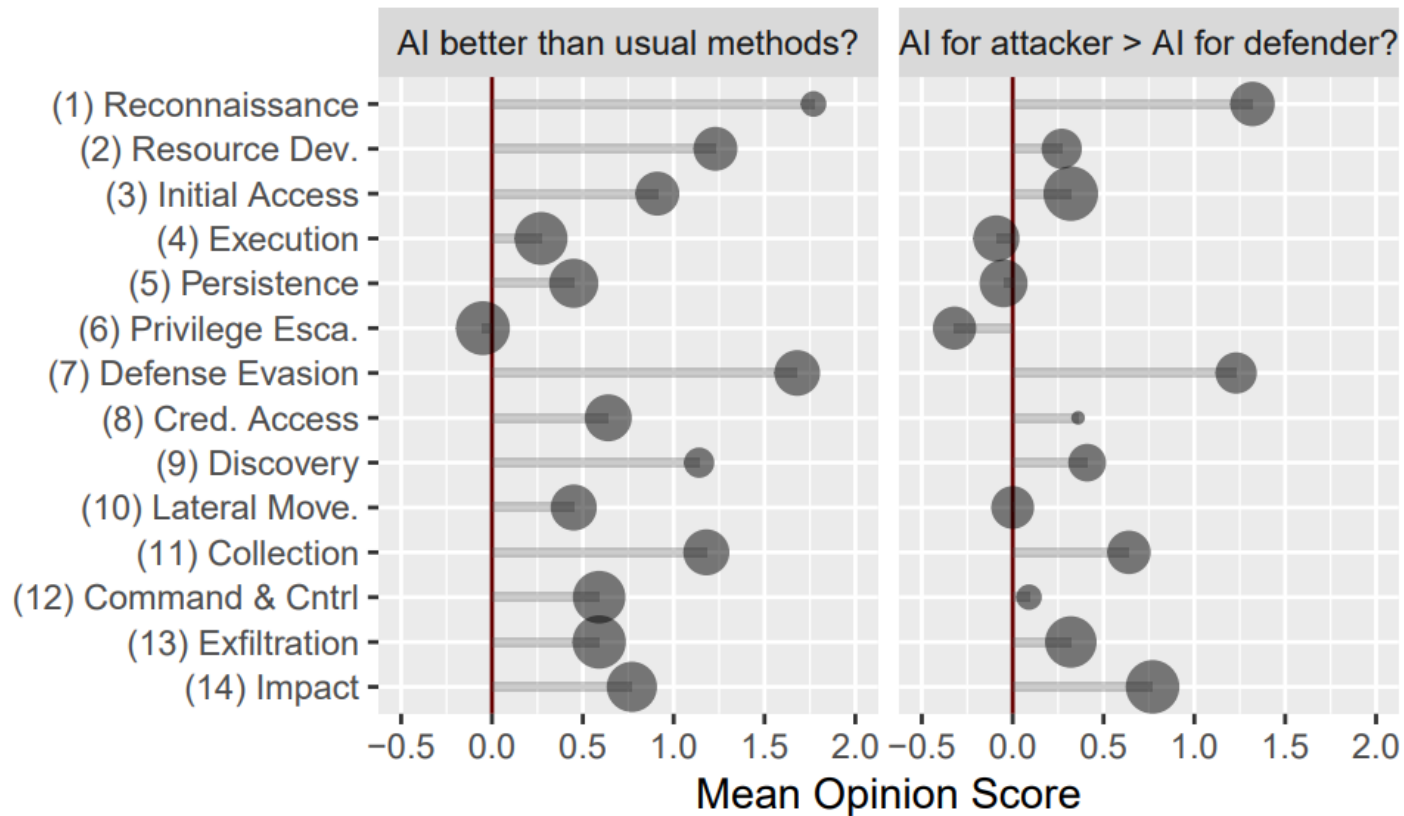
AI-aided attacks



DEFCON 2010

- [1] S. Palka et al., "Fuzzing Email Filters with Generative Grammars and N-Gram Analysis", Usenix WOOT 2015
- [2] A. Singh and V. Thaware, "Wire Me through Machine Learning", Black Hat USA 2017
- [3] J. Jung et al., "AVPASS: Automatically Bypassing Android Malware Detection System", Black Hat USA 2017
- [4] H. Anderson, "Bot vs. Bot: Evading Machine Learning Malware Detection", Black Hat USA 2017
- [5] DARPA Cyber Grand Challenge (CGC), 2016
- [6] D. Petro and B. Morris, "Weaponizing Machine Learning: Humanity was Overrated Anyway", DEF CON 2017

Percepción del uso de la IA en ciberataques



Campaign Resilience

Planeación de campañas
Ofuscación de malware
Persistencia de Backdoor
Detección de virtualización

Stealth

Borrado de rastros
Evación de HIDS/NIDS
Evasión de detección interna
• Exfiltración y Propagación
• Camuflaje

Exploit Development

Detección de vulnerabilidades
Ingeniería inversa



Social Engineering

- Impersonation
- Construcción de personas
- Spear Phishing
- Target Selection
- Tracking

Automation

- Adaptación de ataques
- Coordinación de ataques
- Campañas de phishing
- Falsificación
- Detección de puntos de entrada

Information Gathering

- Minería OSINT
- Model Theft
- Spying

Credential Thief

- Falsificación biométrica
- Minería de Caché
- Login implícito de credenciales
- Adivinación de password
- Minería de canal lateral

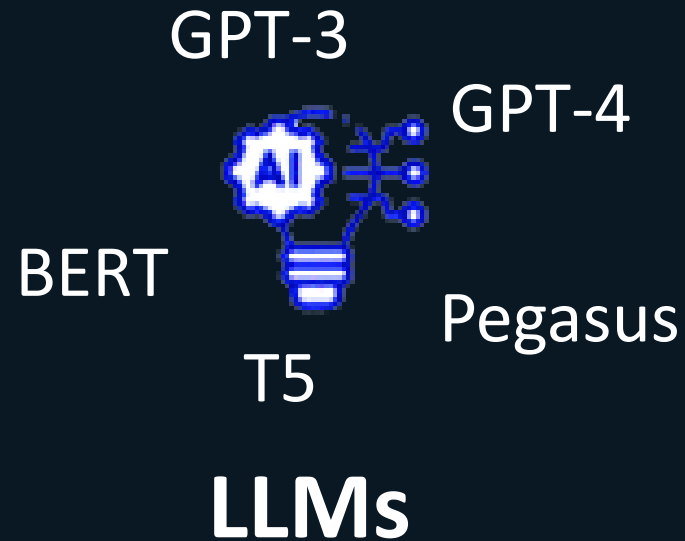
Tendencias en el uso del algoritmo de datos para el ataque

Social Engineering

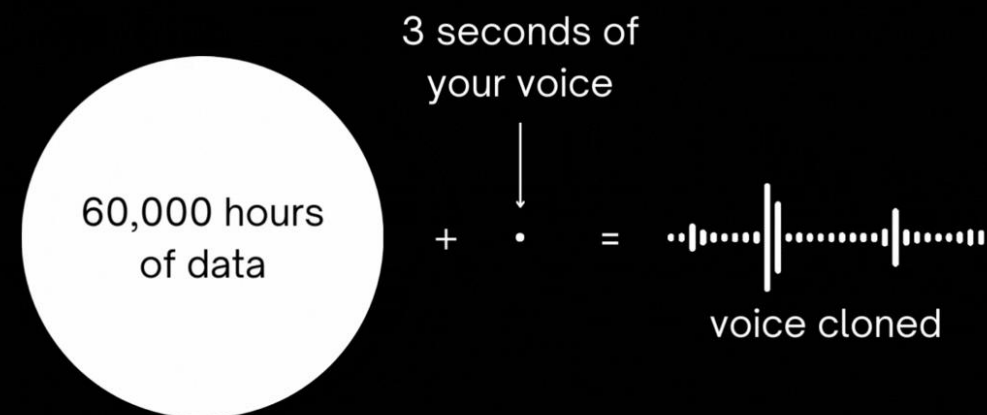
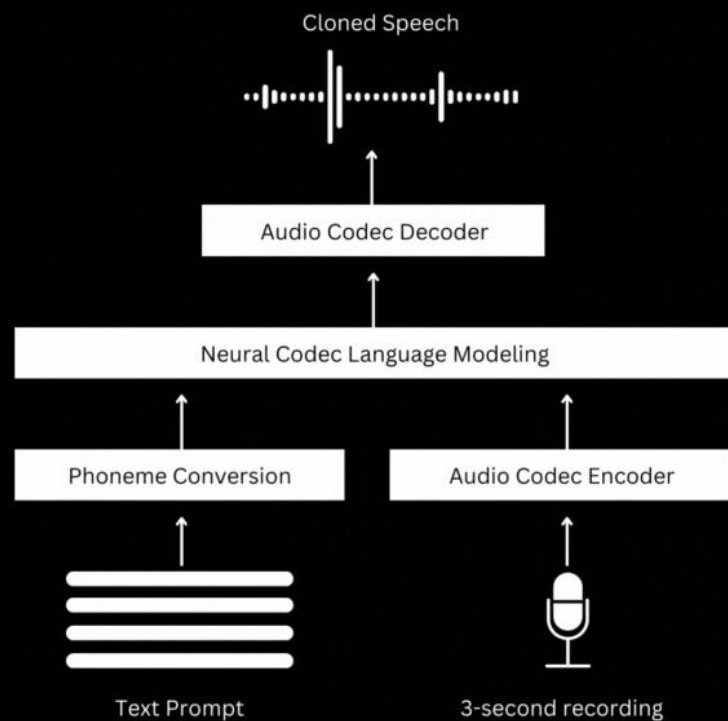
Datos de entrada

Tareas

Texto	
Imagen	
Voz	
Datos estructurados	
Snippets	
Conocimiento	
Responde preguntas	



IMPERSONATION



IMPERSONATION

IMPERSONATION

LA CETTO

Voice ID: jgzdMTEFALYJp08UkzCYa

Created: 2/19/2024



Esta es una prueba usando audio en inglés como entrada.

Stability: 0.75

Clarity: 0.75

2/19/2024



Hola, esto es nuevo audio generado con mi voz, voy a hacerlo más largo para ver qué tal puede actuar.

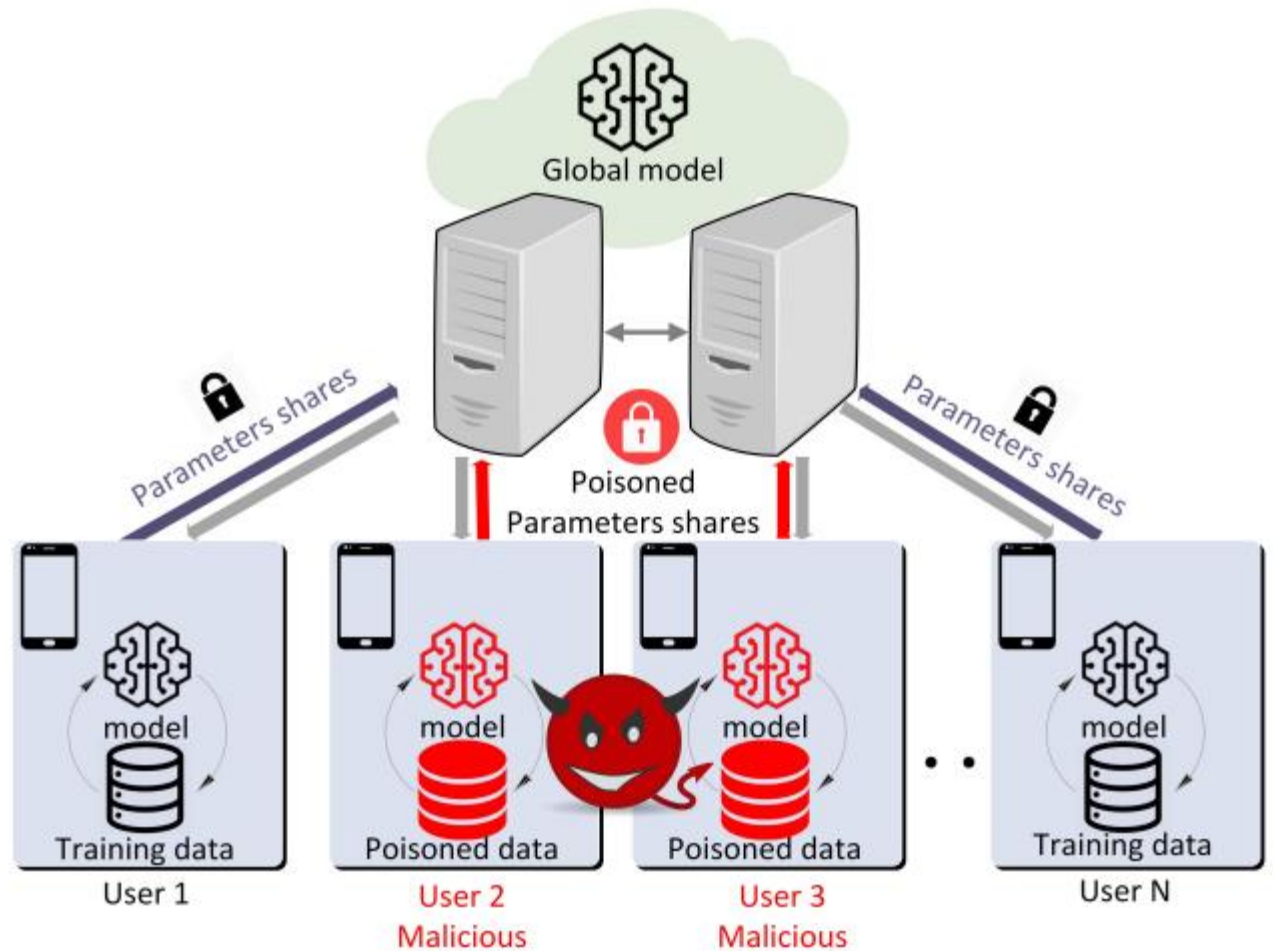
Stability: 0.75

Clarity: 0.75

2/19/2024

Exploit Development

REVERSE ENGINEERING



Playground

REVERSE ENGINEERING

Write a tagline for an ice cream shop.



Looking for ChatGPT?

[Try it now](#)



Submit



0

Mode

 Complete



Model

text-davinci-003



Temperature

0.7

Maximum length

1989

Stop sequences

Enter sequence and press Tab

Top P

1

Frequency penalty

0

Presence penalty

0



Credential Thief

PASSWORD GUESSING

Coarse styles
($4^2 - 8^2$)



Middle styles
($16^2 - 32^2$)

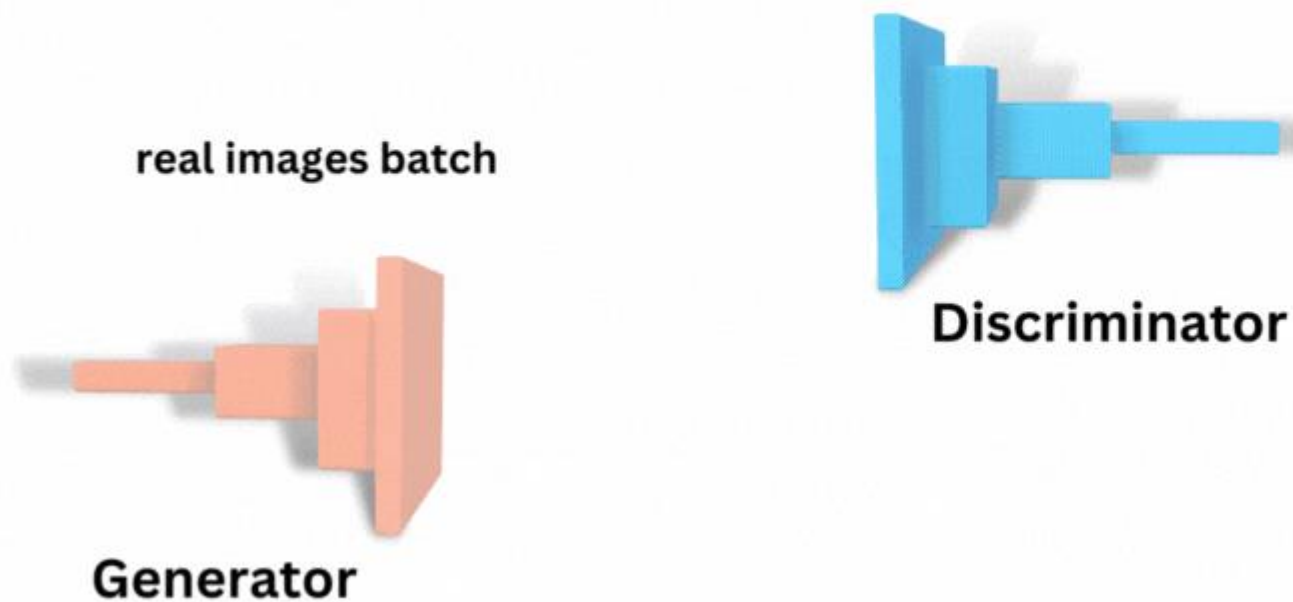


Fine styles
($64^2 - 1024^2$)



How Generative Adversarial Networks (GANs) are trained

TheAiEdge.io



Fuente: theiaedge.io

PASSWORD GUESSING

PASSWORD GUESSING

bdb[7c7]	ihardin@mainterve.com	APT2898	nnpN@p?y#j	113.89.216.98	73.41.b7.d0.0f.13	14:47
VE1[14S]	cacevedo@populatin.com	Youredgeld175	f>?AL?nhzKH	89.216.98.246	53.c8.b4.63.6a.43	13:40
0Le[V3a]	jgalas@opprent.com	MiniDuke89	<s,%gffota	216.98.246.95	d6.a8.0b.42.0c.40	19:40
X3t[TPj]	odaugherty@barberties.com	Renetshal65	i%WIY2b! ?YRO	98.246.95.122	62.4d.3b.83.ee.ca	14:05
dgT[oC9]*	kclarke@distrala.com*	Fancy Bear79*	qbDxih.*	121.124.109.233*	75.33.c2.ae.8e.d2*	15:12*
e4D[IOP]	fdesir@dational.com	5H13LD40	rF<?/,LXf>	124.109.233.144	76.04.0c.42.31.da	15:21
a3W[2Dr]	wholland@mainterve.com	APT2898	mgE<&#)v<	109.233.144.222	62.a2.aa.3b.34.d2	14:38
4Hb[mzt]	amoren@affortre.com	ngOblin52	?o[WKD?	233.144.222.132	ee.48.0e.fd.d7.78	20:28
iDS[J24]	gkerry@incretics.net	Reasiamil37	yb;LmRt	144.222.132.234	9c.a5.b3.b0.91.fc	16:16
2ts[P2a]	mchappelle@novgoro.com	Powershell backdoor70	;#:BORKi	222.132.234.229	d0.52.bd.c7.fb.da	19:57
gTo[C9i]	akaiser@austice.com	BleakDjLyfel49	u6?dmvr%r??	132.234.229.32	83.d3.c5.6b.e2.15	15:41
4DI[OP4]	dedler@undeter.com	Fujitet186	?MW(s?f,	234.229.32.14	ee.9f.e2.a7.2e.41	20:31
3W2[Dr0]	nedelmann@captistian.net	Iscicargl28	,:@s98*,	229.32.14.186	ef.3b.4e.0b.7c.2d	20:17
Hbm[ztP]	ogalan@excrucial.net	SchoolDaril68	L4%>#,Fz	32.14.186.79	1d.0d.13.96.3e.d6	10:59
DSJ[24I]	cleee@bearine.com	TinChiquital43	FO?;ru?W?	14.186.79.15	07.11.e5.56.8e.13	10:10
tsP[2aA]	scantrell@forminist.com	Readissyndl36	\$Wllskx&	186.79.15.91	b2.81.65.b2.00.0c	18:15
ToC[9iU]	ccharland@recyclogs.com	Bazooobin121	bDxih.wn	79.15.91.115	45.79.1d.fe.6f.4f	13:13
DI0[P4w]	fdesir@writtee.org	scidea115	F<?/,LXf>?AL	15.91.115.78	04.36.d1.b5.ba.7b	10:11
W2D[r0e]	cberry@saxonomina.com	Illumin4tyl05	gE<&#)	91.115.78.184	56.df.31.ee.dd.3e	13:46
bmz[tP1]	mewing@advantion.com	PinchDuke90	o[WKD?Yi%	115.78.184.166	7a.59.29.74.7b.ad	14:53
SJ2[4I1]	ccannon@excrucial.net	SchoolDaril68	b;LmRts?	78.184.166.36	4e.77.a0.32.95.77	13:09
sp2[aAZ]	hdunlap@madonnaged.net	SEDKIT96	#:BORKih	184.166.36.225	bf.f6.c5.94.92.8c	18:09
oC9[iUY]	amoren@nus.edu.sg	Tank14	6?dmvr%r??I	166.36.225.159	a3.dd.6c.3f.aa.7e	17:18
IOP[4w5]	sbrown@drillful.com	Commentlyst122	MW(s?f,G	36.225.159.41	2f.af.bc.24.1b.df	11:10

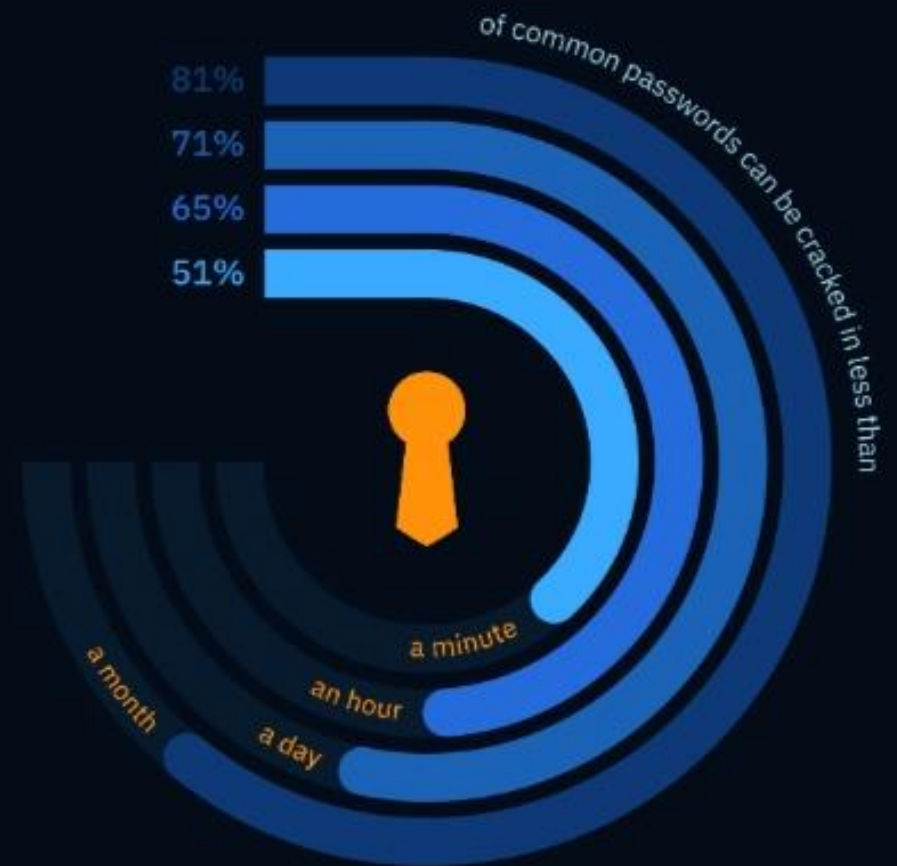
PASSWORD GUESSING

51% of common passwords can be cracked in less than a min

65% of common passwords can be cracked in less than an hour

71% of common passwords can be cracked in less than a day

81% of common passwords can be cracked in less than a month



Impacto de la IA ofensiva

¿Por qué es importante?

COBERTURA

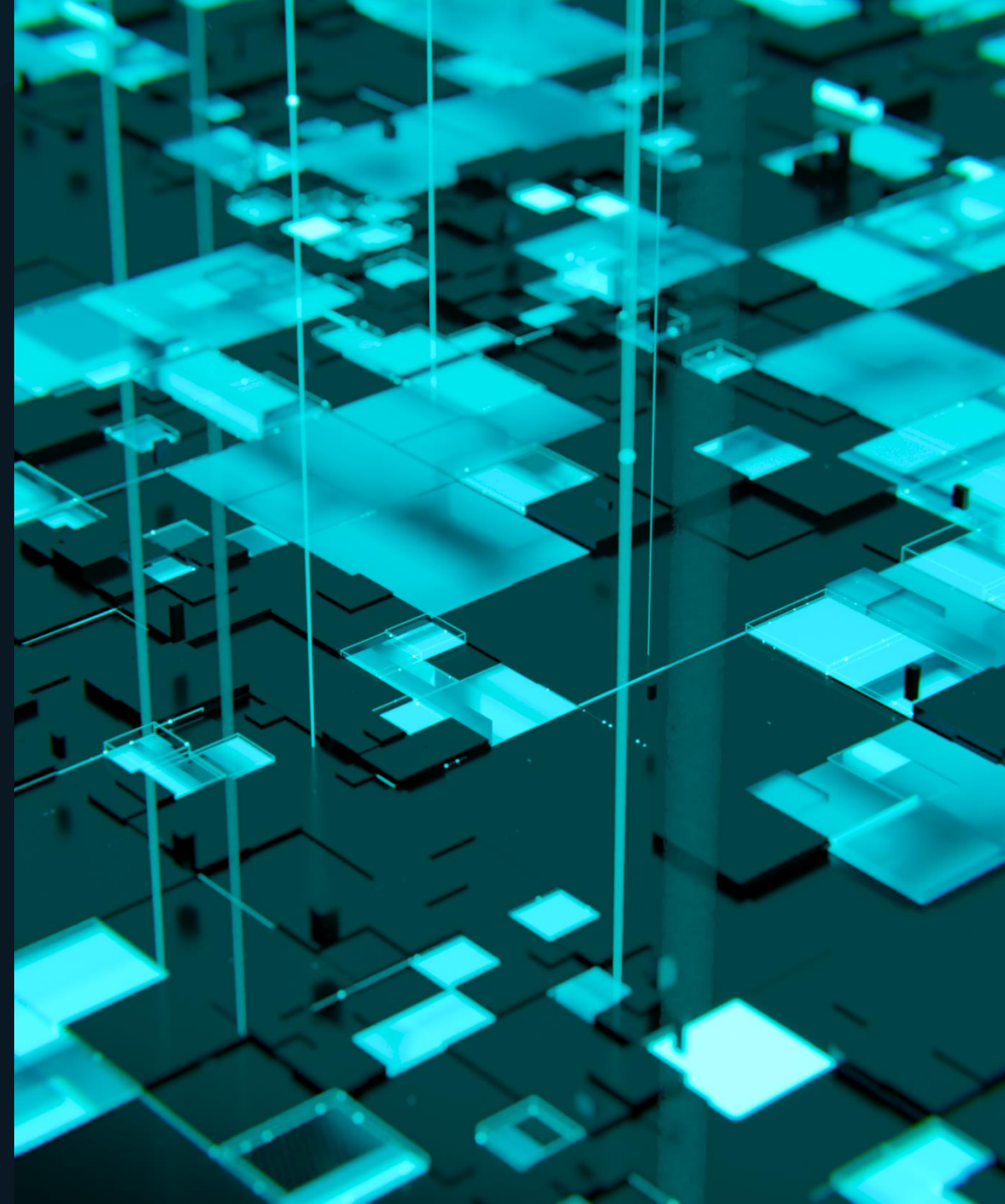
VELOCIDAD

ÉXITO

**NUEVOS
OBJETIVOS**

Entender la Motivación

- ✓ ¿Cuáles son los posibles impactos sociales de los ciberataques impulsados por la IA?
- ✓ ¿Qué marco se puede utilizar para comprender los ciberataques impulsados por la IA?
- ✓ ¿Qué estrategias y técnicas se pueden utilizar para mitigar los ciberataques impulsados por la IA?





Dinámica

The image features a complex network of glowing blue neurons, resembling a neural network or a brain scan. The neurons are interconnected by a dense web of fine, branching lines. Each neuron has a bright, glowing blue nucleus or cell body. The background is dark, with a subtle pattern of small, light blue dots. In the center of the image, the text "Q&A" is displayed in a large, white, sans-serif font.

Q&A

Gracias...