

# Aspectos legales de la seguridad informática

Mikel Egaña Aranguren

[mikel-egana-aranguren.github.io](https://mikel-egana-aranguren.github.io)

[mikel.egana@ehu.eus](mailto:mikel.egana@ehu.eus)



# Aspectos legales de la seguridad informática

<https://doi.org/10.5281/zenodo.4302267>

<https://github.com/mikel-egana-aranguren/EHU-SGSSI-01>



# Delitos informáticos

Convenio sobre ciberdelincuencia de la Unión Europea (2001)

- Delitos relacionados con el contenido
- Delitos relacionados con las infracciones de derechos de autor
- Delitos relacionados con la informática
- Delitos contra la confidencialidad, integridad y disponibilidad de datos y sistemas informáticos

# Delitos relacionados con el contenido

- Pornografía infantil
- Amenazas
- Calumnias
- Difusión de contenidos racistas y xenófobos (Por ejemplo en Alemania contenido nazi o negación del holocausto)

# Calumnia vs injuria

Calumnia: decir que alguien ha cometido un delito

Injuria: atentar contra la dignidad de alguien

# Delitos relacionados con las infracciones de derechos de autor

- Propiedad intelectual e industrial
- Distribución de copias ilegales de películas y canciones
- Reproducción de programas informáticos protegidos

# Delitos relacionados con la informática

Falsificación informática que produzca la alteración, borrado o supresión de datos informáticos que ocasionen datos no auténticos: fraudes, estafas, tráfico de contraseñas, etc.

# Delitos contra la confidencialidad, integridad y disponibilidad de datos y sistemas informáticos

- Acceso ilícito a sistemas informáticos (Delitos contra la intimidad, revelación de secretos de empresa, uso no autorizado de equipos informáticos)
- Interceptación ilícita de datos
- Interferencia de los datos que provoquen daños
- Distribución de virus

# Código penal Español

Ley orgánica 10/1995

- Delitos contra la intimidad y el secreto de las comunicaciones (art. 197.1)
- Estafas electrónicas (art. 248.2)
- Infracción de los derechos de propiedad intelectual (art. 270)
- Delitos de daños (art. 264.2)
- Utilización de ordenadores y de terminales de telecomunicaciones sin consentimiento de su titular (art. 256)

# Código penal Español

Ley orgánica 10/1995

- Descubrimiento y revelación de secretos contenidos en documentos o soportes informáticos (art. 278)
- Falsedad en documentos electrónicos (art. 390)
- Fabricación o tenencia de útiles para la comisión de delitos (art. 400)
- Distribución entre menores de edad de material pornográfico (art. 186)

# Código penal Español

Ley orgánica 10/1995

- Distribución de pornografía infantil (art. 189)
- Publicación de calumnias o injurias

# Protección de datos personales

Derecho a la intimidad y privacidad: derecho de poder excluir a terceros del conocimiento de su vida personal (Sentimientos, datos biográficos, imagen, ...)

# Protección de datos personales

Ley orgánica 15/1999, de 13 de diciembre: LOPD

Declaración de derechos humanos 1948: nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques

# Ley Organica Proteccion de Datos de Carácter Personal (LOPD)

Se aplica a organizaciones públicas, privadas y profesionales independientes que almacenen datos personales para tratamiento, uso o explotación posterior (Con exenciones como el censo electoral o los boletines oficiales)

# LOPD: Responsable de Fichero

- Documento de seguridad: actualizado a las leyes vigentes en materia de seguridad de los datos personales
- Adoptar las medidas necesarias para que el personal conozca las normas en materia de seguridad y consecuencias de su incumplimiento
- Implantar un mecanismo de identificación de usuarios
- Mantener una relación de los usuarios del sistema con los derechos de acceso a los datos y aplicaciones

# LOPD: Responsable de Fichero

- Establecer mecanismos para evitar que los usuarios accedan a recursos con derechos distintos de los autorizados
- Verificar los procedimientos de copia y de recuperación de datos
- Autorizar por escrito la ejecución de procedimientos de recuperación de datos

# LOPD: Responsable de Fichero

- Autorizar expresamente el tratamiento fuera de los locales de la organizaciones
- Autorizar la salida de soportes informáticos fuera de los locales de la organizaciones
- Adoptar las medidas correctoras de las deficiencias detectadas en las auditorías de seguridad

# LOPD: Principios

- **Habeas data** ([Habeas corpus](#)): los datos pertenecen al usuario, no a la organización que los almacena
- Calidad de los datos: los datos tienen que ser pertinentes al fin para el que han sido recogidos, y solo conservados durante el tiempo necesario para ese fin
- Seguridad de los Datos
- Deber de secreto incluso después de haber finalizado la relación
- Información en la recopilación de los datos

# LOPD: Principios

- Consentimiento del afectado para el tratamiento
- Cesión de datos solo con el consentimiento
- Cesión a terceros países: solo a países con el mismo nivel de protección
- Datos especialmente protegidos: salud, ideología, vida sexual, origen racial, religión o creencias

# LOPD: Derechos

- Derecho de información en la recopilación de datos
- Derecho de consulta al Registro General de Protección de Datos
- Derecho de acceso a sus datos de carácter personal
- Derecho de rectificación y cancelación
- Derecho de oposición
- Derecho de una indemnización

# Agencia Vasca de Protección de Datos

<https://www.avpd.euskadi.eus/>

# Reglamento General de Protección de Datos (RGPD)

Reglamento Europeo para unificar los reglamentos de todos los países

25 Mayo 2016

# RGPD: nuevos principios

Principio de responsabilidad (accountability): implementar mecanismos que permitan acreditar que se han adoptando todas las medidas necesarias (responsabilidad proactiva)

Protección de datos por defecto y desde el diseño

Principio de transparencia

# RGPD: nuevas obligaciones para entidades

- Designar un Delegado de Protección de Datos (DPO)
- En algunos casos evaluaciones de impacto sobre la privacidad
- Las empresas multinacionales sólo tendrán como interlocutora a una sola autoridad de control (ventanilla única)
- Las brechas de seguridad deberán ser comunicadas a la autoridad de control en menos de 72h y a los usuarios en brechas graves
- Más datos sensibles: genéticos, biométricos, condenas penales

# RGPD: nuevas obligaciones para entidades

- Garantías adicionales para las transferencias internacionales de datos
- Sellos y acreditaciones para la responsabilidad proactiva
- Desaparece la obligación de inscribir los ficheros
- Sanciones más sustanciales, que pueden llegar a los 20 millones de euros o el 20% de la facturación de una empresa

# RGPD: nuevos derechos para ciudadanos

- Mayor transparencia y información
- Consentimiento inequívoco, libre y revocable, mediante acto afirmativo claro  
(No se admite consentimiento tácito)
- Derecho al olvido
- Derecho a la limitación temporal del tratamiento
- Portabilidad de los datos

# Factor humano

Mikel Egaña Aranguren

[mikel-egana-aranguren.github.io](https://mikel-egana-aranguren.github.io)

[mikel.egana@ehu.eus](mailto:mikel.egana@ehu.eus)



# Factor humano

<https://doi.org/10.5281/zenodo.4302267>

<https://github.com/mikel-egana-aranguren/EHU-SGSSI-01>



# El factor humano

“

*Al final, un sistema de seguridad es tan efectivo como lo es el más débil de sus eslabones. En el caso de la seguridad online, el eslabón más débil es siempre el factor humano*

**Eugene Kaspersky**

# El factor humano

“

*Usted puede tener la mejor tecnología, firewalls, sistemas de detección de ataques, dispositivos biométricos, etc. Lo único que se necesita es una llamada a un empleado desprevenido y acceden al sistema sin más.*

**Kevin Mitnick**

# El factor humano

Kevin Mitnick en los 90's fue considerado el Cybercriminal más buscado por el FBI

En uno de sus primeros ataques de ingeniería social explicaba cómo necesitaba un número de solicitante para "pinchar" el Departamento de Vehículos de Motor (DMV)

# El factor humano

Para lograrlo llamó a una comisaría y se hizo pasar por alguien del DMV. "¿Su código de solicitante es el 36472?", a lo cual el agente contestó: "No, es el 62883"

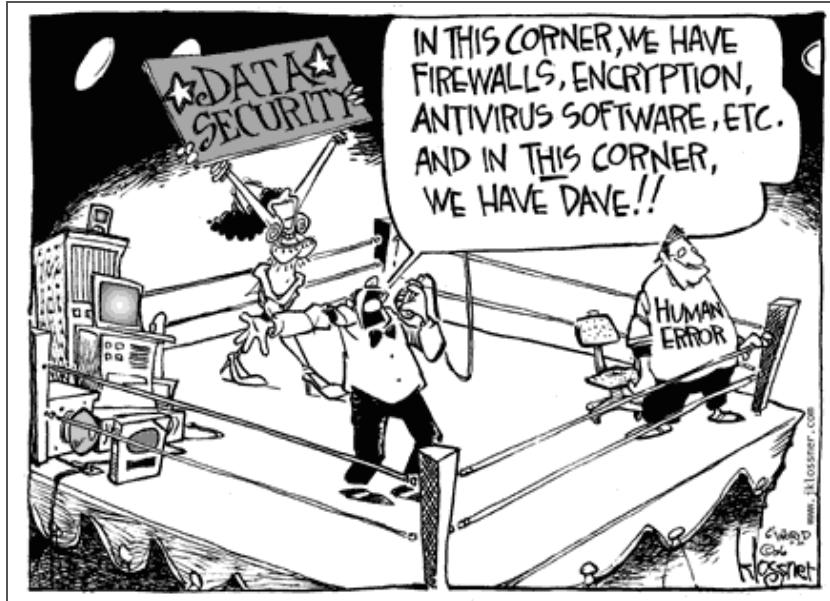
*Es un truco que he descubierto que funciona muy a menudo. Si pides información confidencial, la gente, sospecha de inmediato*

# El factor humano

*Si finges que ya tienes esa información y dices algo que está mal, la gente suele corregirte y te recompensa con la información que estabas buscando*

*Ese principio básico de la ingeniería social se unía a otro esencial: la gente suele ser el eslabón más débil de una cadena de seguridad, porque "la gente siempre tiene intención de ayudar"*

# El factor humano



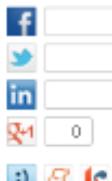
# El factor humano

## Un asesor de Obama, 'cazado' en Facebook

- Jon Favreau pide disculpas a Hillary Clinton por difundir en Internet una fiesta con una silueta de la ex primera dama

EFE / ELPAÍS.COM | 6 DIC 2008 - 01:20 CET

Archivado en: Estados Unidos · Tecnología



Una de las principales características de la campaña del presidente electo de Estados Unidos, **Barack Obama**, que ha llevado al máximo partido a las redes de contacto social, particularmente a Facebook, ha traído problemas a uno de sus principales asesores.

El próximo director de Discursos de la Casa Blanca, el escritor y guionista Jon Favreau (derecha), ha sido acusado de haber difundido en Internet una foto en la que aparece bebiendo alcohol y besando a una figura que se asemeja a la ex primera dama Hillary Clinton.



El asesor Jon Favreau (derecha) aparece junto a una figura de Hillary Clinton.

# El factor humano

POLÉMICA EN LA RED

## Paula Vázquez la lía en Twitter

La popular presentadora publica por error en internet su número de teléfono móvil

22.10.12 - 19:00 - REDACCIÓN | MADRID

0 Comentarios | [Twittear](#)

Compartir

Recomendar

110



Conectado a diariovasco.disqus.com...

<http://www.diariovasco.com/rc/20121022/mas-actualidad/tecnologia/paula-vazquez-twitter-201210221857.html>

# El factor humano

**VIRALES** 09/02/2018 11:08 CET | Actualizado 09/02/2018 11:09 CET

## Rosalía publica por error el número de teléfono de Pablo Alborán en Instagram

Se ha marcado un Paula Vázquez.

[https://www.huffingtonpost.es/2018/02/09/rosalia-publica-por-error-el-numero-de-telefono-pablo-alboran-en-instagram\\_a\\_23357228/](https://www.huffingtonpost.es/2018/02/09/rosalia-publica-por-error-el-numero-de-telefono-pablo-alboran-en-instagram_a_23357228/)

# El factor humano

Tweets

Fátima Báñez García @FatimaBanez  
¡Obtuve 5390 puntos en Bubble Shooter Adventures! ¿Puedes mejorararlo? ghh [goo.gl/S44Cb](http://goo.gl/S44Cb) [pic.twitter.com/P48LDY49](http://pic.twitter.com/P48LDY49)

 Ocultar aplicación    Responder    Retwittear    Favorito

desarrollado por  Photobucket    Reporta este archivo

# El factor humano

## Cosidó, pillado jugando en horas de trabajo

El SUP denuncia que el director general de la Policía se dedica a jugar por Internet mientras que los policías "tienen que ir a trabajar estando enfermos"

Estrella Digital, @Estrella\_digi. 12/06/2013 | 10:26 h.

0 comentarios



**Ignacio Cosidó** @Ignacos

He volado 170m en un juego repleto de acción de Jetpack Joyride.  
¡Supera eso! [bit.ly/rKuWqK](http://bit.ly/rKuWqK) [pic.twitter.com/EwuXWd2Sz3](http://pic.twitter.com/EwuXWd2Sz3)

[View photo](#)

6m

# El factor humano

EN ACTITUD CARIÑOSA

## Eduardo Casanova (Fidel en 'Aída') cuelga accidentalmente una imagen en internet practicando sexo con su novio

El actor aparece frente al espejo desnudo junto a su pareja. 26 Septiembre 2012.



Los peligros de la red se hacen más latentes para los famosos. [Eduardo Casanova](#) puede dar fe

<http://www.formulatv.com/noticias/27106/eduardo-casanova-fidel-aida-cuelga-accidentalmente-imagen-sexo-novio/>

# El factor humano

## El presidente de Nuevas Generaciones del PP en Huesca se burla de la violencia machista

■ José Luis Ferrando tuiteó una imagen en la que una joven narcotizada es amordazada y arrastrada por un hombre con el texto "¡he ligado!"

eldiario.es Seguir a @eldiariօs 61 comentarios

04/10/2013 - 18:59h Me gusta 12.17 Twitter 1.51

Tweet

J.L. Ferrando Castro @JL\_Ferrando Yujuuuuuuu pic.twitter.com/i6UgjxkndP

8:05 AM - 15 sep 13 desde Huesca, Huesca



# El factor humano

## CONSEJO DE SEGURIDAD EN EL USO DEL CORREO ELECTRÓNICO

Los problemas que hemos tenido este último mes para el envío de correos se deben a que algunos usuarios han facilitado su usuario y contraseña a spammers. Por ello, desde la vicegerencia TIC queremos hacer las siguientes aclaraciones:

1.- **NUNCA LE PEDIREMOS SU USUARIO Y CONTRASEÑA** por correo electrónico. **NUNCA**.

Por tanto, cualquier mensaje que reciba en el que se le solicite, no ha sido enviado por nosotros y por tanto debe usted tratarlo como una falsificación.

2.- **NUNCA DEBE ENVIAR SU USUARIO Y CONTRASEÑA POR CORREO ELECTRÓNICO**, ni a nosotros ni a otra persona. **NUNCA**. No es el medio indicado para hacer esto.

En caso de que los necesitemos para hacer alguna prueba, no se los pediremos por correo electrónico.

3.- Los mensajes que envía esta vicegerencia se suelen enviar en castellano y euskera, y en todo caso con una sintaxis correcta. Si recibe un mensaje con muy mala sintaxis, desconfíe de él.

4.- Ante la menor duda sobre un mensaje de este estilo, descártelo. Si necesita aclaraciones, póngase en contacto con el CAU y solicítelas, siempre antes de responder.

<http://www.ehu.es/correow>

# El factor humano

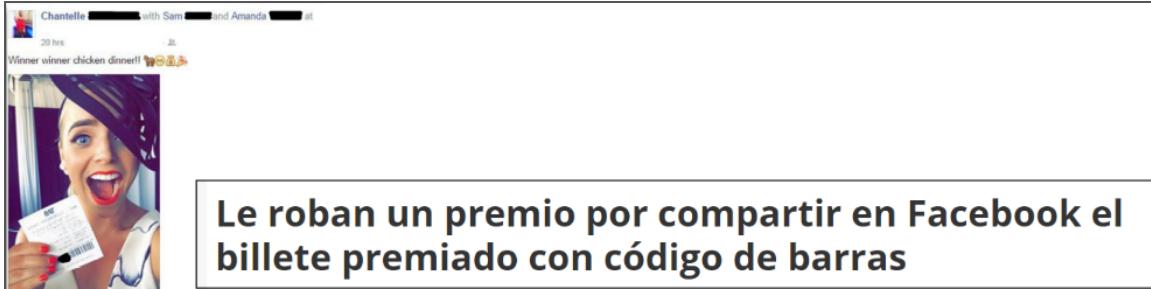
Un tuit racista provoca el despido fulminante de una directiva en pleno vuelo

Justine Sacco escribió "Me voy a África. Espero no pillar el sida. Es broma. ¡Soy blanca!" e inició una tormenta en Twitter que acabó con su carrera profesional

Tecnología | 23/12/2013 - 17:46h | Última actualización: 24/12/2013 - 17:48h

<http://www.lavanguardia.com/tecnologia/20131223/54397498289/un-tuit-racista-provoca-el-despido-fulminante-de-una-directiva-en-pleno-vuelo.html>

# El factor humano



Chantelle [REDACTED] with Sam [REDACTED] and Amanda [REDACTED] at [REDACTED]  
20 hrs · 20 likes · 2 comments  
Winner winner chicken dinner!! 🎉🐔🍗

**Le roban un premio por compartir en Facebook el billete premiado con código de barras**

[https://www.abc.es/recreo/abci-roban-todas-pertenencias-publicar-foto-facebook-201608081854\\_noticia.html](https://www.abc.es/recreo/abci-roban-todas-pertenencias-publicar-foto-facebook-201608081854_noticia.html)

# El factor humano

PIRATERÍA INFORMÁTICA ›

## Los altavoces inteligentes pueden recibir órdenes de terceros inaudibles para el usuario

El fallo es una puerta para que los 'hackers' puedan actuar sobre unos dispositivos que cada vez son más populares

[https://elpais.com/tecnologia/2018/05/11/actualidad/1526030082\\_845494.html](https://elpais.com/tecnologia/2018/05/11/actualidad/1526030082_845494.html)

# El factor humano

**Strava: cómo una aplicación de deportes dejó al descubierto secretos de bases militares de Estados Unidos**

Redacción  
BBC Mundo

© 29 enero 2018

f t e m Compartir



<https://www.bbc.com/mundo/noticias-42859883>

# El factor humano

Los usuarios también son parte del sistema

- También generan problemas de seguridad (Involuntarios o intencionados)
- Hay que tenerlos en cuenta en las políticas de seguridad
- Detrás del éxito de una gran parte de los ataques informáticos se encuentra un usuario “inocente”

# El factor humano

¿Cómo son los ataques intencionados?

- El 75% de las empresas temen represalias de ex empleados
- Robo de información
- Sabotaje

# El factor humano

¿Cómo se evitan los ataques intencionados?

- No siempre se puede, sobre todo a priori (¿Cómo distinguir si la intención es buena o mala?)
- Ante las dudas, auditorías

# El factor humano

Las empresas deberían

- Evaluar los riesgos
- Evaluar su exposición a los mismos
- Preparar una respuesta por si se producen

# El factor humano

A nivel preventivo

- Acceso limitado a los datos
- Medidas “extra” de seguridad para datos importantes

# El factor humano

¿Cómo se aprovechan del factor humano los hackers/crackers?

- Desconocimiento / Ignorancia
- Dejadez / Pereza
- Curiosidad / Ganas de saber / Ganas de lucrarse
- Comunicación / Ganas de darse a conocer
- Miedo
- Vergüenza / Desprestigio

# El factor humano

Desconocimiento / Ignorancia

- ¿Cómo se actualiza el sistema operativo?
- ¿Hay que actualizar las aplicaciones?
- Este mensaje de nueva versión de Java que aparece, ¿Qué hago?
- Mejor no toco nada no vaya a dejar de funcionar
- Total, ¿quién va a querer acceder a mi ordenador?
- ¿Necesitas mi password? Apunta, es ...

# El factor humano

Estimado Mikel:

Atendiendo a su solicitud para usuario en WebUntis, le comunico sus datos:

Usuario: [REDACTED]

Contraseña: [REDACTED]

Saludos cordiales,

[REDACTED]

Administratiboa

Administrativo

[REDACTED]



Bilboko Ingenieritzaz Eskola Escuela de Ingeniería de Bilbao  
Euskal Herriko Unibertsitatea Universidad del País Vasco

Plaza Ingeniero Torres Quevedo, 1. 48013 Bilbao

[www.ehu.eus](http://www.ehu.eus)



# El factor humano

Dejadez / Pereza

- De 3,4 millones de claves de 4 dígitos filtradas
- 11% de las claves eran 1234
- 6% de las claves eran 1111
- 2% de las claves eran 0000

# El factor humano

Dejadez / Pereza

- 100,000 passwords de trabajadores de Apple, Google, Nasa, etc. en el IEEE
- 271 trabajadores tenían 123456
- Más de 200 tenían ieee2012 (año de la filtración)
- Más de 200 tenían 12345678

# El factor humano

Dejadez / Pereza

- Cambiar el password cada 6 meses es muy pesado
- Memorizar un password seguro para cada aplicación es muy pesado
- Instalar 21 actualizaciones de Windows... uff! Con la prisa que tengo

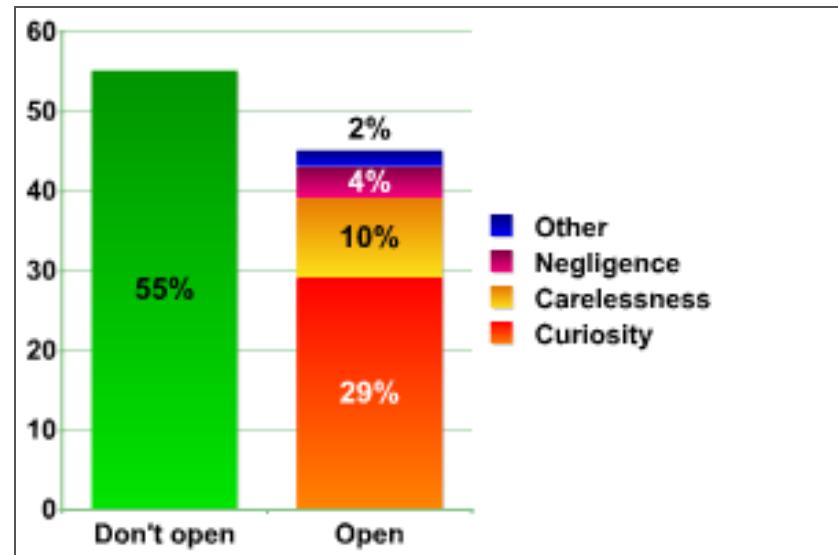
# El factor humano

Curiosidad / Ganas de saber / Ganas de lucrarse

- Mira la foto de la fiesta ...
- Han pillado a esta pareja en actitud cariñosa.. No te lo pierdas!!
- ¿Quieres un puesto de trabajo?
- Medicinas online
- Te ha tocado la lotería de Sudáfrica!!
- Tengo una herencia que no puedo cobrar, ¿lo haces tú y te llevas una comisión?

# El factor humano

Curiosidad / Ganas de saber: ¿Qué hacen los usuarios ante un correo no deseado?



# El factor humano

Comunicación / Ganas de darse a conocer: Típico en redes sociales

- Me voy de vacaciones!!
- Fotos, gustos, datos personales
- ¿Todos tus "amigos" / seguidores son amigos tuyos? ¿Les conoces personalmente? ¿Confías en ellos?
- ¿Quién tiene acceso a tu información?

# El factor humano

Vergüenza / Desprestigio

- Las personas no denuncian por vergüenza
- Las empresas no denuncian por desprestigio
- Consecuencia: Los timadores siguen lucrándose

# El factor humano

¿Cómo se aprovechan del factor humano los hackers/crackers? Ingeniería Social

- Obtener información confidencial a través de un usuario
  - De manera pasiva (sin interactuar con él)
  - A través de redes sociales
  - Seguimientos
- Se engaña al usuario para que proporcione información (técnicas activas)

# Ingeniería Social

La información obtenida de forma pasiva se puede utilizar para muchas cosas:

- Intentos de encontrar contraseñas: fechas/nombres significativos, aficiones, ...
- Para usarla luego en un ataque:
  - Correo fraudulento del banco
  - Conocimiento sobre el objetivo en general

# Ingeniería Social. Técnicas

Scam:

- Estafa a través de correo electrónico o páginas web
- Puede haber pérdida económica o no
- Hoax, phishing, spam, pharming

# Ingeniería Social. Técnicas

Hoax:

- Intento de hacer creer que algo falso es real
- No suelen tener consecuencias económicas
- Generan tráfico inútil y sobrecargan servidores
- Peligro: el cuento de Pedro y el lobo (Cuando algo sea real, el usuario no se lo creerá)
- Juegan con los miedos / buena intención de los usuarios

# Ingeniería Social. Técnicas

Hoax (Prevención):

- Suelen ser anónimos y no citan fuentes
- Contienen una petición de reenvío
- Pensar con lógica
- No reenviar / publicar aquello que no estamos completamente seguros que es real. En caso de duda, informarse

# Ingeniería Social. Técnicas

Phishing:

- Intento de lograr contraseñas o datos bancarios a través de un correo o una web que aparenta ser oficial
- Suele usarse en conjunto con el envío de SPAM
- El enlace muestra una cosa y redirige a otra
- URL muy parecida a la original: <http://www.kutzabank.es/>
- URL con mismo nombre, pero distinto dominio: <http://www.bankia.bz/>

# Ingeniería Social. Técnicas

Técnicas de Phishing:

- Cross Site Scripting (inyectar código malicioso en la página real)
- IDN Spoofing (vulnerabilidad en nombres de dominio internacionales por el uso de Unicode). Los navegadores actualizados no son vulnerables

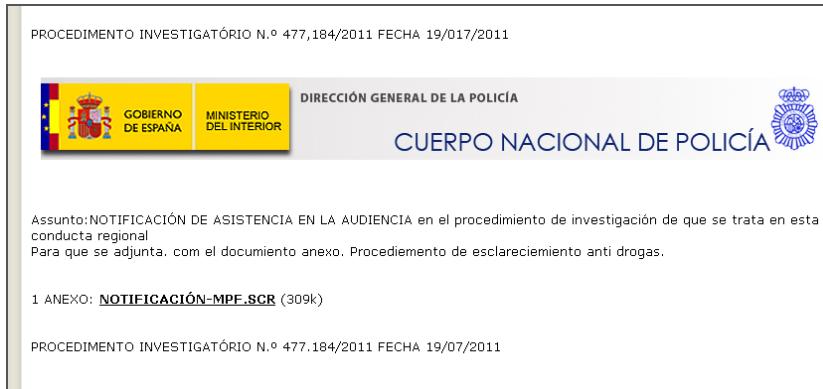
# Ingeniería Social. Técnicas

Phishing:

- Los ataques suelen ser masivos
- **Spear Phishing:** Ataques dirigidos a objetivos concretos

# Ingeniería Social. Técnicas

Correo con fichero adjunto que infecta el ordenador y "roba" información



# Ingeniería Social. Técnicas

Introduce tus datos para recibir la devolución de la Renta

 Agencia Tributaria

### Forma de Reembolso

Avisos:

1. Por favor, introduzca sus datos personales y una tarjeta de crédito válida a la que desea efectuar la devolución.
2. Todos los campos son obligatorios.

Nombre Completo:

Fecha de Nacimiento:  - Dia -  - Mes -  - Año -

Dirección:

Ciudad:

Código Postal:

Número de Tarjeta:

Fecha de Caducidad:  - Mes -  - Año -

Código de Seguridad:

Cantidad a devolver:  223.56 EUR

# Ingeniería Social. Técnicas

## Aplicaciones de redes sociales



# Ingeniería Social. Técnicas

Soluciones al Phishing:

- Nunca dar información confidencial por e-mail
- Teclear directamente la dirección, no pinchar un enlace
- Comprobar que la conexión esté cifrada (HTTPS)
- Comprobar los certificados

# Ingeniería Social. Técnicas

Soluciones al Phishing:

- Usar versiones actualizadas de los navegadores
- Usar un antivirus que analice las webs que se visitan
- Usar un servicio de análisis de URLs

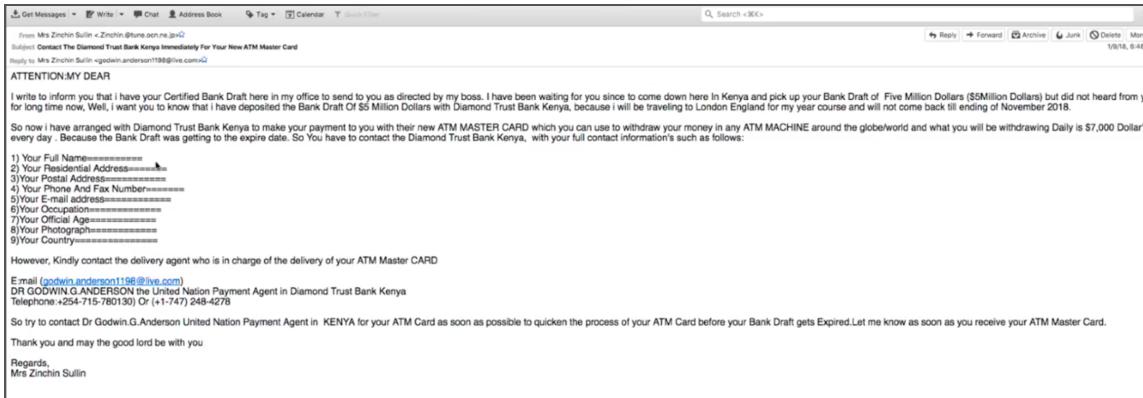
# Ingeniería Social. Técnicas

Timo nigeriano (estafa 419):

- Teclear directamente la dirección, no pinchar un enlace
- Se usa en conjunto con el SPAM
- Herencias, loterías, posibles parejas, ...

# Ingeniería Social. Técnicas

## Ejemplo variante timo nigeriano



# Ingeniería Social. Técnicas

## Ejemplo variante timo nigeriano

The screenshot shows an email inbox interface with various menu options at the top: Get Messages, Write, Chat, Address Book, Tag, Calendar, and Quick Filter. Below the menu, an incoming email is displayed.

**From:** Mrs.Melania Trump <WWW@festa.ocn.ne.jp>  
**Subject:** First notice from Mrs.Melania Trump.  
**Reply to:** Mrs.Melania Trump <melaniatrump777@gmail.com>

**First notice from Mrs.Melania Trump.**  
I am Mrs Melania Trump and I am written to inform you about your Bank Check Draft brought by United Embassy from the government of Benin Republic to the white house Washington DC and has been mandated to be deliver to your home address,as soon as you get back to me with your below information.

1.Full Names :  
2.Residential Address :  
3.Mobile Number:  
4.Fax Number :  
5.Occupation :  
6.Sex :  
7.Age :  
8.Nationality :  
9.Country :  
10.Marital Status :

Your check is containing the sum of \$25 million USD.  
Here is my email or send me an sms,+1(407) 990-1723 but i prefer sms because ill always busy in the white house and i cant be able to

# Ingeniería Social. Técnicas

Soluciones al timo nigeriano:

- Pensar antes de actuar
  - Nadie regala dinero
  - Si no se juega a la lotería, es imposible que toque
- No dar información confidencial a desconocidos

# Ingeniería Social. Técnicas

## Herencias

Estimado amigo,

Soy Emmanuel Egobiawa, un abogado en derecho y abogado personal para fines Ingeniero S. García, que murió con su esposa y su único hijo en un accidente de coche espantoso en el día 13 de diciembre de 2008, que utilizan para trabajar en la Compañía de Desarrollo de Shell y También era un contratista del gobierno aquí en Lomé. Deseo llamar su atención para informarle que Engr tarde. S. García antes de su muerte dejó a la suma de dieciocho millones de dólares (EE.UU. \$ 18,000,000, 00) solo en su cuenta bancaria que quiero poner en su atención ahora. Él murió sin dejar ninguno de sus familiares la información a mí o a cualquier otra persona y tengo mis mejores tratar de localizar a sus parientes o familiares, incluso en la embajada de su país, pero sin ningún éxito. Ahora bien, como su abogado personal y por la ley y el orden, el banco me pedirá que proporcione a sus familiares o parientes más cercanos a este hombre para que el fondo / el dinero se traslado a su familia que no tienen.

Ahora ya no tiene ningún miembro de la familia o parientes como (familiares hermano, hermana, tío o familiar), y tener / respuesta el mismo apellido (García) con él, quiero y han decidido a presentar al banco como uno de sus miembros de la familia o pariente más cercano a él por lo tanto ponerse en contacto con usted para que el banco va a transferir este dinero / fondos en su cuenta. Después de recibir este fondo / dinero en su cuenta en su país, voy a venir a su país a efectos de compartir y de la inversión porque parte de este fondo / el dinero se debe utilizar para la Fundación del Orfanato y otras inversiones como la construcción de una buena Estate en su país que se nos está dando otro fondo adicional / dinero. Pero esto no se puede lograr sin un socio extranjero como a ayudar a mí llevar a cabo esta operación, y que es por eso que estoy en contacto con usted hoy en día para que me ayude en este tema. Tengo los documentos necesario para que nos ayude en la toma de este éxito.

# Ingeniería Social. Técnicas

## Loterías



The National Lottery®

Premio Asegurado

PO Box 251 Watford WD18 9BR  
Inglaterra.

24<sup>th</sup> junio 2011.

Desde: International Award Dept.  
Reference Number: WB/2011/0018  
Batch Number: BC-00067/5808

Attention: Beneficiario

PREMIO ASEGURADO

Tenemos el immense placer de informarle hoy día 08 de Abril 2011, el resultado de las promociones de loterías "UK NATIONAL LOTTERY". llevado a cabo el dia 22 de Abril 2011.

Su nombre con su email ha sido premiado adjunto al boleto: 026-9-2 con número de serie: 7-8 mostró el número afortunado De Remesa: 1-8-3. En consecuencia, ganador de la lotería en tercera categoría. Por lo tanto, a usted le ha correspondido un premio de €915.000,00 euros (NOVECIENTOS QUINCE MIL EUROS) en efectivo. El número de referencia de archivo para reclamar su premio es: GTC1/2551256003/09. El premio total en efectivo es €19.733.910 euros (DIECINUEVE MILLONES SETECIENTOS TREINTA Y TRES MIL NOVECIENTOS DIEZ EUROS). Compartido entre varios ganadores a diferente escala internacional en esta categoría 3. Felicitaciones!

Todos los participantes han sido seleccionados a través de un sistema informático, llevado a cabo anualmente. En este momento, su dinero se encuentra depositado en una cuenta provisoria a su nombre, bajo un seguro que nuestra empresa ha puesto a su dinero para tenerlo asegurado. Para mayor seguridad, le pedimos que guarde bien esta documentación, ya que aquí figura su número de referencia y cualquier persona que posea estos datos podría reclamar el dinero en su nombre.

Para comenzar su demanda, debe ponerse en contacto con el número de teléfono que aquí le indicamos, y su agente le informara el procedimiento para el cobro correspondiente a su dinero. Teléfono: +44 [REDACTED] Email: [REDACTED]@firstsecurity.com FIRST SECURITY COMPANY LTD Persona responsable de asesoramiento: ALAMS DOUGLAS. Horario comercial: Lunes a Viernes de 10 a 14 hs y de 17 a 20 hs. NOTA: Todo premio debe ser reclamado antes de 26 de Julio de 2011. Despues de esta fecha, los fondos serán devueltos al MINISTERIO DE ECONOMIA Y HACIENDA como no reclamado.

RELLENE EL FORMULARIO Y ENVIARLO POR E-MAIL AL TU AGENCIAS JUNTO CON TU PHOTOCOPIA DE TU DNI EMAIL: [REDACTED]@firstsecurity.com

# Ingeniería Social. Técnicas

## Trabajo (Muchas veces ilegal)

**Asunto:** Trabajar en casa, pago semanal de 1.768 euros por semana.

Bienvenida.

**Aumentamos nuestra dependencia y necesitamos le..**

Si no esta satisfecho con sus ingresos- aprovechar la oportunidad para convertirse en remoto te propuesto nuestro corporacion y cobrar de 10 a 30 euros por hora en la Internet.

Todo lo que necesita- posesion nivel de usuario de PC, disponibilidad y una demanda enviada,  
que contengan datos de nombre completo, edad y lugar de residencia.

**Encuesta que desea expulsar aqui [www@west-ug.org](mailto:www@west-ug.org)**

Ya un par de horas. Le enviaremos una carta en respuesta con explicaciones de la obra detalladas.

**Solo esperamos de usted responsabilidad y el deseo para ganar. Y ningunos costes iniciales!**

# Ingeniería Social. Técnicas

## Regalos



# Ingeniería Social. Técnicas

Para detectar SPAM, revisar la cabecera:

- From -- el remitente
- To -- El destinatario
- Subject -- El asunto del mail
- Date -- La fecha de envío
- **Received** -- Indica en cada línea por qué servidores ha pasado (en orden inverso) -- Se puede usar el Servicio [Whois](#)

# DMARC

DMARC (Domain-based Message Authentication, Reporting & Conformance) es un estándar que autentica el dominio del remitente de correos electrónicos, para que tanto los remitentes como los destinatarios puedan verificar los mensajes entrantes

Se definen las medidas que deben aplicarse a los mensajes sospechosos que se reciban

# DMARC

Comprobaciones de DMARC:

- Los mensajes entrantes deben estar autenticados por SPF, DKIM o ambos
- El dominio autenticado debe concordar con el que figura en la dirección del encabezado "De:" del mensaje

# Spoofing de correo electrónico

- Spoofing (suplantación): cambiar el contenido de un mensaje, para que parezca que proviene de una fuente que no es la real
- Los spammers pueden enviar correos electrónicos de modo que parezca que proceden de tu dominio

# DKIM (Domain Keys Identified MaiL)

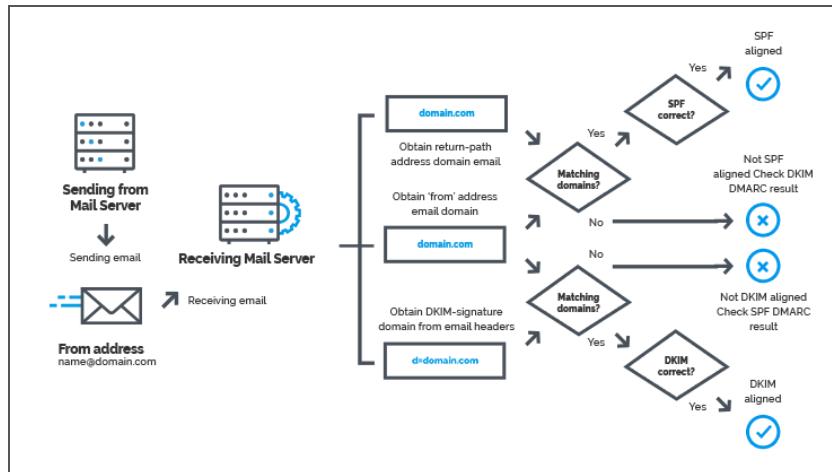
- DKIM previene más fácilmente el spoofing en los mensajes salientes que se envíen desde tu dominio
- DKIM incluye una firma cifrada en el encabezado de todos los mensajes salientes: Los servidores de correo electrónico que los reciben descifran su encabezado mediante DKIM y verifican que no se haya modificado tras el envío

# SPF (Sender Policy Framework)

- Protegerse frente a correos falsificados que parecen proceder de tu dominio

# DMARC

Google, Facebook, Microsoft, etc. están bloqueando el phishing y ataques de spam con DMARC



# Ejemplo real Media Markt

Gmail lo clasifica como spam

The screenshot shows an email from MediaMarkt in the inbox. The subject line is: "Y para el fin de semana... ACER y ROWENTA ¡2ª unidad de la misma marca al -50%! + solo hasta el 26/10 LG, XIAOMI, OPPO, VSMART y ORAL B". A "Spam" button is visible next to the message. Below the message, there's a tooltip asking why it's in spam and a "No es spam" button. The message content includes a link to "¡2ª unidad al 50% de la misma marca!". At the bottom, there are links for "Contacto", "Social Media", and "Formas de pago". A context menu is open over the message, showing options: "Responder", "Reenviar", "Filtrar mensajes como este", "Imprimir", "Eliminar este mensaje", "Bloquear a MediaMarkt", "Denunciar suplantación de identidad", "Mostrar original" (which is highlighted), "Descargar mensaje", and "Marcar como no leído".

# Ejemplo real Media Markt

## MX ToolBox Email Head Analyzer

**Header Analyzed**  
Email Subject: 📌 Y para el fin de semana... ACER y ROWENTA 📌 | 2ª unidad de la misma marca al -50%! + solo hasta el 26/10 LG, XIAOMI, OPPO, VSMART y ORAL B

**Delivery Information**

- > DMARC Compliant
- > SPF Alignment
- > DKIM Unauthenticated
- > DKIM Alignment
- > DKIM Authenticated

**Relay Information**

Received	Delay
923 seconds	Delay:

**Hop**   **Delay**   **From**   **By**   **With**   **Time (UTC)**   **Blacklist**

Hop	Delay	From	By	With	Time (UTC)	Blacklist
1	.	uspmta194148.emarsys.net 217.175.194.148	mx.google.com	ESMTPS	10/25/2019 10:08:54 PM	●
2	0 seconds		2002:a2e:9c12:0:0:0:0	SMTP	10/25/2019 10:08:54 PM	
3	15 minutes		2002:a92:6c09::	POP3	10/25/2019 10:24:16 PM	
4	1 Second		2002:a05:6214:8f:0:0:0	SMTP	10/25/2019 10:24:17 PM	

**SPF and DKIM Information**

# Black list

blacklist:217.175.194.148 [Monitor This](#) [Solve Email Delivery Problems](#) [blacklist](#)

! We notice you are on a blacklist. [Click here for some suggestions](#)

Checking 217.175.194.148 against 99 known blacklists...  
Listed 2 times with 3 timeouts

	Blacklist	Reason	TTL	ResponseTime	
<span>✗</span> LISTED	SORBS SPAM	217.175.194.148 was listed <a href="#">Detail</a>	3600	0	<a href="#">Ignore</a>
<span>✗</span> LISTED	UCEPROTECTL2	217.175.194.148 was listed <a href="#">Detail</a>	2100	0	<a href="#">Ignore</a>
<span>✓</span> OK	0SPAM			0	
<span>✓</span> OK	Abuse.ro			142	
<span>✓</span> OK	Abusix Mail Intelligence Blacklist			0	

# Black list

**SORBS** (Spam and Open Relay Blocking System) proporciona acceso a la lista negra antispam

**UCEPROTECTL2** (Unsolicited Commercial E-mail). Las listas negras (mala reputación) basadas en spam son aquellas que enumeran direcciones IP individuales o rangos completos, del que se han recibido spam. Por ejemplo, correo electrónico masivo no solicitado

# Ingeniería Social. Pharming

Redireccionar el tráfico de una web legítima a otra falsa

- Atacando el servidor DNS
- Atacando el fichero hosts en local

Peligroso porque el usuario ha introducido correctamente la URL: El redireccionamiento es “invisible”. Prevención:

- Si el aspecto de la web es diferente, sospechar
- Comprobar certificados

# Ingeniería Social

La única forma de luchar contra la Ingeniería Social

- Educación de los usuarios
- Implementación de políticas de seguridad que realmente se sigan

Cuanta más información nuestra tengan los timadores, más fácil será que nos engañen

# Casos reales. El director general chulito

Auditoría de seguridad para una compañía

El director general alardea de su seguridad

El consultor descubre los donativos a instituciones de lucha contra el cáncer

# Casos reales. El director general chulito

A través de Facebook se descubre el restaurante y el equipo deportivo favoritos del director general

Llama al director general haciéndose pasar por una de las asociaciones de lucha contra el cáncer con las que colabora habitualmente

A cambio de la donación entra en sorteos de cenas en su restaurante favorito y entradas para su equipo favorito

El director general accede a recibir más información por correo electrónico

# Casos reales. El director general chulito

Para asegurarse que no va a haber problemas al abrir el fichero, se le pregunta al director qué versión de Adobe Reader usa

Se le envía un fichero .pdf con código malicioso para esa versión concreta

Se consigue acceso total al ordenador del director general y desde ahí a toda la empresa

# Casos reales. El parque temático

Contratan a una consultora para analizar la seguridad de sus sistema de venta de entradas

El consultor llamó al parque temático haciéndose pasar por vendedor de software

Tras hablar un rato con los empleados obtuvo la información de qué versión de Adobe Reader se usaba en el parque

# Casos reales. El parque temático

El consultor se presenta en el parque simulando una familia (con niños)

Pide acceso a un ordenador para poder imprimir las entradas que tiene en el correo electrónico

La empleada le permite el acceso (a pesar de tenerlo prohibido)

# Casos reales. El parque temático

Al abrir el archivo .pdf con las entradas, se instala un software malicioso que permite controlar el ordenador

Desde ese ordenador se accede a los servidores de la empresa

# Informática forense

Mikel Egaña Aranguren

[mikel-egana-aranguren.github.io](https://mikel-egana-aranguren.github.io)

[mikel.egana@ehu.eus](mailto:mikel.egana@ehu.eus)



# Informática forense

<https://doi.org/10.5281/zenodo.4302267>

<https://github.com/mikel-egana-aranguren/EHU-SGSSI-01>



# ¿Qué es la Informática forense?

Disciplina criminalística

Investigar sistemas informáticos para obtener y procesar información (evidencias digitales):

- Con validez jurídica
- Para la simple investigación privada (accesos no autorizados, sospechas de robos de información, etc.)

# ¿Qué es la Informática forense?

Trata de responder:

- ¿Qué?
- ¿Quién?
- ¿Cómo?
- ¿Cuándo?
- ¿Por qué?

# ¿Qué es la Informática forense?

Es utilizada por:

- Agentes de la ley
- Compañías de seguros
- Compañías privadas
- Personas particulares
- ...

# ¿Qué es la Informática forense?

Consiste en:

- Extraer información de un sistema
- Recuperar información cifrada/eliminada/dañada
- Monitorizar el comportamiento de un sistema
- Detectar incumplimientos de las políticas de la empresa
- ...

# ¿Qué es la Informática forense?

Principio de intercambio de Locard:

- "Siempre que dos objetos entran en contacto transfieren parte del material que incorporan al otro objeto"
- Todas las acciones dejan un rastro

# ¿Qué es la Informática forense?

Principio de incertidumbre de Heisenberg:

- "El mero hecho de medir el estado de un sistema lo altera"
- No se puede obtener información de un sistema sin modificar el sistema
- Obtener la mayor cantidad de información posible minimizando las alteraciones y su impacto

# ¿Qué es la Informática forense?

La validez jurídica de una evidencia digital la decide el juez

Todo documento, log, máquina, etc. ha podido ser manipulado/accedido por terceros

¿Un documento con una firma electrónica reconocida tiene validez jurídica?

# ¿Qué es la Informática forense?

... ¿Y si el acusado alega que le robaron el certificado (la tarjeta)? ¿Hay denuncia? ¿Se solicitó la revocación del certificado inmediatamente?

# ¿Qué es la Informática forense?

Para que las evidencias digitales tengan validez jurídica hay que seguir procesos que aseguren:

- Que se ha respetado la ley para obtenerlas
- Que la información es exactamente la que se recogió
- Que durante su análisis no se ha modificado/creado/eliminado nada
- El análisis realizado tiene que poder ser reproducible

# ¿Qué es la Informática forense?

Forensic Examination of Digital Evidence: A Guide for Law Enforcement

Electronic Crime Scene Investigation: A Guide for First Responders, Second  
Edition

UNE 71506 - Metodología para el análisis forense de las evidencias  
electrónicas

# ¿Qué es la Informática forense?

Good Practice Guide for Computer-Based Electronic Evidence

RFC 3227 - Guidelines for Evidence Collection and Archiving

ISO/IEC 27037:2012 Information technology -- Security techniques –  
Guidelines for identification, collection, acquisition and preservation of  
digital evidence

# Informática forense. Proceso

1. Identificación
2. Conservación
3. Análisis
4. Exposición

# Informática forense. Proceso

Es imprescindible tomar notas, grabaciones, fotografías, vídeos, etc. de todo lo que se realiza con fechas y horas

Puede ser necesario recordar todo el proceso con la mayor cantidad de detalles posibles en un juicio (años después)

# Identificación

Identificar los sistemas (evidencias) que van a ser necesarios en la investigación

Es aconsejable la presencia de un notario que de fe de todo lo que se realiza

Conviene tomar fotografías que muestren su disposición/configuración

# Identificación

Desde el primer momento hay que activar la cadena de custodia: registrar de manera exhaustiva quién maneja las evidencias recogidas indicando fechas, horas, dónde se almacenan, quien es el responsable de su custodia, etc.

Si son sistemas que están en marcha, evitar que se sigan usando y recoger toda la información volátil (Podría borrarse al apagar el sistema): Usar programas externos para realizar copias, accesos, etc.

# Identificación

La información de la memoria RAM es muy importante (Hay que copiarla modificándola lo menos posible):

- Procesos en ejecución
- Módulos y DLL's en ejecución
- Archivos abiertos
- Claves del registro abiertas
- Versiones desencriptadas de datos

# Identificación

La información de la memoria RAM es muy importante (Hay que copiarla modificándola lo menos posible):

- Adjuntos de Email, imágenes, fragmentos de chat
- Llaves criptográficas
- Contraseñas en texto plano
- ...

# Identificación

Herramientas para volcar el contenido de la RAM:

- pd Process Dumper
- FTK Imager
- Volatility
- EnCase

# Identificación

Habrá que recoger también la información sobre los procesos en marcha, los servicios, los usuarios conectados a la máquina, los puertos abiertos, etc.

Cuidado!, Si no hay un notario que de fe de qué se ha hecho y de la información que se ha obtenido... ¿Quién asegura que eso era exactamente lo que había en el sistema en ese momento?

# Identificación

Una vez recogida toda la información volátil se apaga el sistema y se copia toda la información no volátil (Discos duros, USBs, etc.)

Es conveniente el uso de Write Blockers, sistemas que permiten acceder a la información, pero evitan la escritura en el disco

# Identificación

Se hace una copia bit a bit: Duplicado forense (Así se copian los “restos” y la información oculta que haya por el disco duro)

Se calcula (y se almacena) el resumen criptográfico del original y de la copia para asegurar que son idénticos

Se realiza otro duplicado forense de la copia, para evitar tener que trabajar con el original en caso de daño de la copia

# Identificación

Herramientas de clonado bit a bit:

- dd (comando Linux)
- Helix3 Pro
- EnCase
- FTK Imager

# Conservación

Se deben evitar (Cadena de custodia):

- Pérdidas
- Contaminación
- Daño, alteración, manipulación

# Conservación

Documentar exhaustivamente toda la información recogida

Etiquetar todos los dispositivos recogidos

Indicar marca, modelo, número de serie, etc.

# Conservación

Fecha, datos y firma de las personas que lo trasladen y manipulen

El original debe quedar a buen recaudo (por ej: en poder del notario)

Se puede entregar una copia a todas las partes interesadas

Siempre es aconsejable tener una copia de respaldo

# Análisis

Analizar toda la información obtenida es una tarea tediosa y "casi imposible"

Se usan muchos tipos de herramientas:

- Recuperación de elementos borrados
- Crackeo de passwords
- Analizadores de logs
- ...

Hay que ser ordenado y meticuloso; la intuición del analista es esencial

# Análisis

Sitios típicos de búsqueda de información:

- Correos electrónicos
- Herramientas de mensajería
- Ficheros eliminados
- Metadatos de los ficheros (creación, último acceso, etc.)
- Historiales de navegación
- Logs de aplicaciones y del sistema
- Conexiones a otras máquinas

# Análisis

Es importante manejar la línea temporal del sistema:

- Cuándo se instaló X
- Cuándo se accedió a Y
- Cuándo se borró Z

# Análisis

Es imprescindible respetar la LOPD y el derecho al secreto de las comunicaciones (No se pueden leer correos electrónicos con su médico ni con un amante si no son relevantes para la investigación)

# Análisis

Solución: Búsqueda ciega (Intuición del analista)

- No se examina toda la información
- Se realizan búsquedas por palabras clave
- Sólo se analiza la información donde figuran esas palabras clave

Todo un informe pericial puede ser desestimado si se ha violado alguna ley para realizarlo

# Presentación

Se realiza un informe explicando todo el proceso y los resultados obtenidos

Por muy bueno que haya sido el proceso, y los resultados obtenidos si el informe no lo refleja correctamente, no tendrá valor

El informe está dirigido a personas no técnicas (jueces, abogados, empresarios, etc.). Debe entenderse

El informe debe ser imparcial. El perito no debe expresar opiniones, sólo reflejar pruebas y resultados

# Presentación

Partes de un informe:

- Antecedentes
- Evidencias
- Análisis y tratamiento
- Resultados
- Conclusiones

# Presentación

**Antecedentes:** cuál es la situación que ha hecho necesaria la intervención de un perito

**Evidencias:** evidencias que se han recogido y los procesos que se han seguido de recogida, duplicación, conservación, etc.

**Análisis y tratamiento:** técnicas y herramientas usadas para analizar la información

# Presentación

**Resultados:** se expondrán de modo claro y entendible qué resultados dieron las técnicas empleadas

**Conclusiones:** el apartado más importante. Es en el que el experto explica qué se puede deducir de los resultados obtenidos. Todas las conclusiones tienen que derivarse de algún resultado, si no es una suposición

# Presentación

En el caso de que haya un juicio, el perito actuará en calidad de testigo

Tendrá que explicar el informe que elaboró en su momento y responder a las preguntas de los abogados

Debido a la lentitud de la justicia, han podido pasar varios años. Es conveniente repasar el informe unos días antes del juicio

# Presentación

A veces se llama a declarar a un perito para que desmonte el informe de otro perito:

- Por qué se rompió la cadena de custodia y las evidencias se pudieron alterar
- Por qué las conclusiones del informe no son directamente derivables de los resultados obtenidos
- Por qué aplicando técnicas distintas se obtienen resultados que contradicen los obtenidos en el informe

# Deep Web

Mikel Egaña Aranguren

[mikel-egana-aranguren.github.io](https://mikel-egana-aranguren.github.io)

[mikel.egana@ehu.eus](mailto:mikel.egana@ehu.eus)



# Deep Web

<https://doi.org/10.5281/zenodo.4302267>

<https://github.com/mikel-egana-aranguren/EHU-SGSSI-01>



# ¿Qué es la Deep Web?

La red habitual y conocida (Clearnet):

- Está formada por direcciones conocidas (Ej: [www.ehu.eus](http://www.ehu.eus))
- Con contenidos en HTML que están indexados y permiten realizar búsquedas para encontrar lo que nos interese

# ¿Qué es la Deep Web?

La Deep Web (internet profunda) está formada por todos aquellos contenidos que no son directamente accesibles a través de internet

Se estima que de todo el contenido que existe:

- El 10% está en Clearnet (el internet que conocemos)
- El 90% está en la Deep Web

# ¿Qué es la Deep Web?

Contenido de la Deep Web

- Información confidencial o protegida (No suelen estar indexados por buscadores ni se puede acceder directamente a ellos):
  - Registros sanitarios
  - Registros académicos
  - Datos bancarios
  - ...

# ¿Qué es la Deep Web?

Contenido de la Deep Web

- Información "suelta": por ejemplor un archivo HTML que no esté enlazado desde ningún otro
- Información en formatos no HTML que un navegador no puede leer
- Contenido no publicable (Censura): Contenidos que no pueden publicarse libremente porque pueden acarrear consecuencias

# ¿Qué es la Deep Web?

Contenido de la Deep Web

- Contenido ilegal y/o desagradable (Darknet)
  - Tráfico de armas, drogas, personas
  - Material pedófilo
  - Malware
  - Alquiler de hackers, matones, etc.
  - Películas snuff

# ¿Qué es la Deep Web?

## Niveles de la Web

- Nivel 1: Sitios web ampliamente conocidos y para todos los públicos
- Nivel 2: Sitios web desconocidos y sitios pornográficos
- Nivel 3 (Aquí empieza Deep Web): Necesita privacidad y anonimato (contenidos sensibles)
- Nivel 4: Necesita un proxy. Nivel peligroso (contenidos ilegales)
- Nivel 5: Secretos gubernamentales y militares
- ???

# Formas de acceso

Para acceder a la Deep Web hace falta un software especial que proporcione  
privacidad, anonimato y ejerza de proxy

Existen varias alternativas que darán acceso a distintos contenidos de la  
Deep Web: TOR, I2P, Freenet, Zeronet

# TOR

The Onion Router (TOR)

Red de navegación anónima

Oculta el origen y el destino de los paquetes que navegan por la red

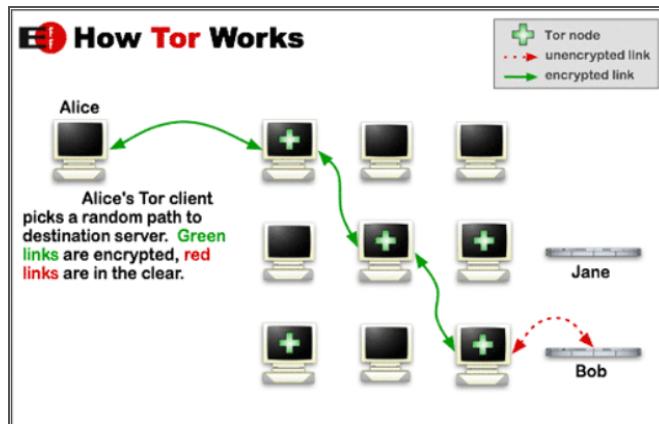
# TOR

Para acceder a TOR se necesita un software específico:

- TOR Browser: navegador web
- Tails (The Amnesic Incognito Live System): Sistema operativo que se ejecuta desde un USB

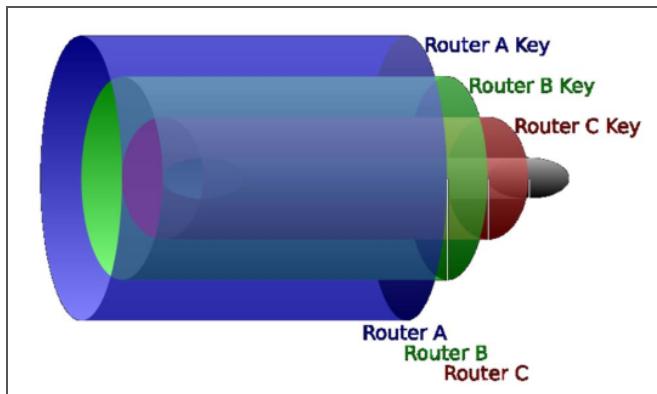
# TOR

Cada vez que hay que hacer una conexión, se calcula un camino aleatorio basado en los nodos de la red



# TOR

La información se cifra a capas (como una cebolla) con las claves públicas de los distintos nodos, de modo que cada nodo sólo puede ver cuál es el siguiente



# TOR

Utilizando la red TOR se puede acceder a URLs que son inaccesibles de otro modo:

- Dominio .onion
- URLs alfanuméricas: <http://3g2upl4pq6kufc4m.onion/>

# TOR

Para encontrar contenidos hay que usar buscadores específicos o sitios donde se recopilen las URLs:

- Buscador Torch (<http://xmh57jrznw6insl.onion/>)
- The Hidden Wiki (<http://kpvz7ki2v5agwt35.onion> )



# **Security and Privacy in Deep Learning.**

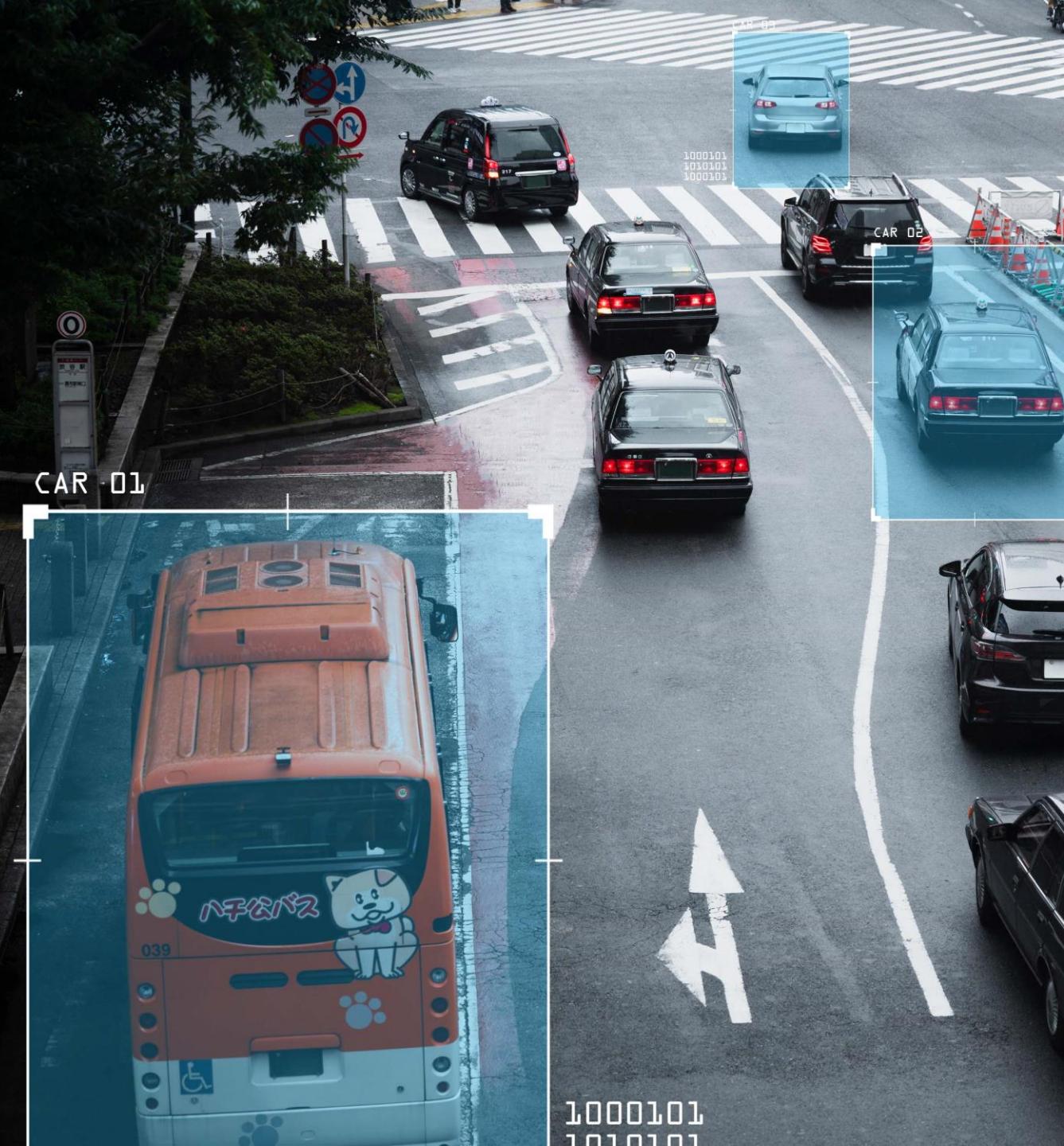
by Gorka Abad



1.

# Introduction.

1. Use cases



# Introduction.

G

Make a brief general introduction to deep learning



Deep learning is a subfield of machine learning that is inspired by the structure and function of the brain, specifically the neural networks that make up the brain. Deep learning algorithms use multiple layers of artificial neural networks to learn and make decisions. These layers process the input data and learn to recognize patterns and features in the data. The layers of the neural network are "deep," hence the name "deep learning."

Deep learning has become very popular in recent years due to the success of deep learning algorithms in a variety of applications, including image and speech recognition, natural language processing, and even playing games like chess and Go. Deep learning algorithms can learn and make decisions on their own, without the need for explicit programming.

There are many different types of deep learning algorithms, including convolutional neural networks, recurrent neural networks, and autoencoders, to name a few. These algorithms are trained on large datasets and can learn to perform a wide range of tasks, such as classifying images, translating languages, or even generating text.

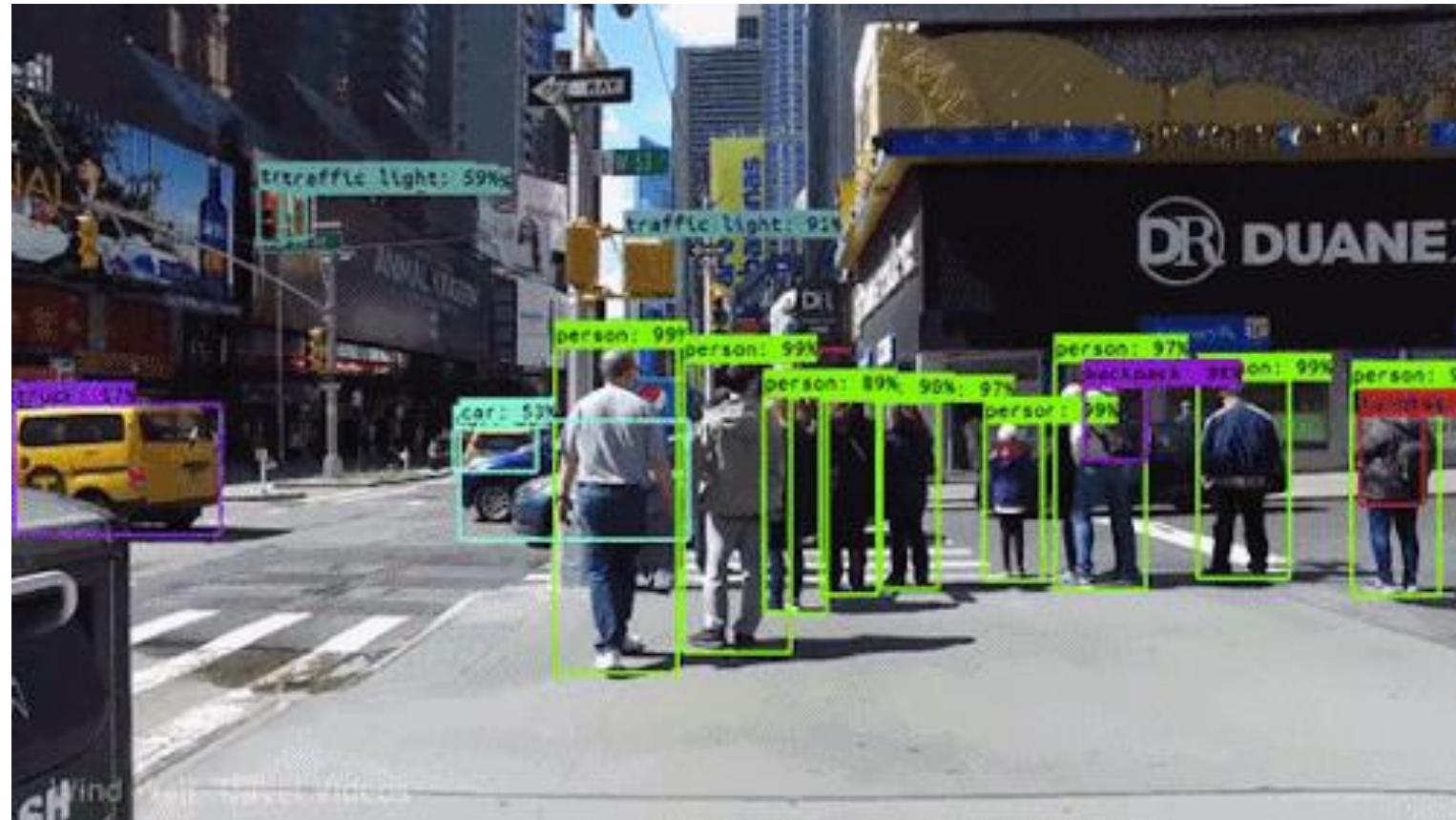
# Introduction.

The slide has a dark background with a light gray header bar. In the top-left corner of the header is a blue square icon containing a white letter 'G'. To its right, the text 'Make it cool' is written in a white sans-serif font. Below the header, there is a green circular icon with a white neural network symbol inside. To its right, the text 'Deep learning is a really cool field of machine learning that has revolutionized many areas of artificial intelligence. It's inspired by the structure and function of the brain, and allows computers to learn and make decisions on their own, without explicit programming.' is displayed in white. Below this text, another line of white text reads 'Deep learning algorithms have been used to achieve some amazing feats, such as:'. A bulleted list follows, each item preceded by a white square icon:

- Image and speech recognition: Deep learning algorithms can recognize and classify images and speech with incredible accuracy, making it possible for computers to understand and respond to human inputs.
- Natural language processing: Deep learning algorithms have made significant strides in understanding and generating human language, enabling applications like language translation and chatbots.
- Playing games: Deep learning algorithms have even been used to play games like chess and Go at a world-class level, demonstrating their ability to learn and adapt to new situations.

At the bottom of the slide, the text 'Overall, deep learning is a really exciting field that is constantly pushing the boundaries of what is possible with artificial intelligence.' is written in white.

# Use cases.



# Use cases.

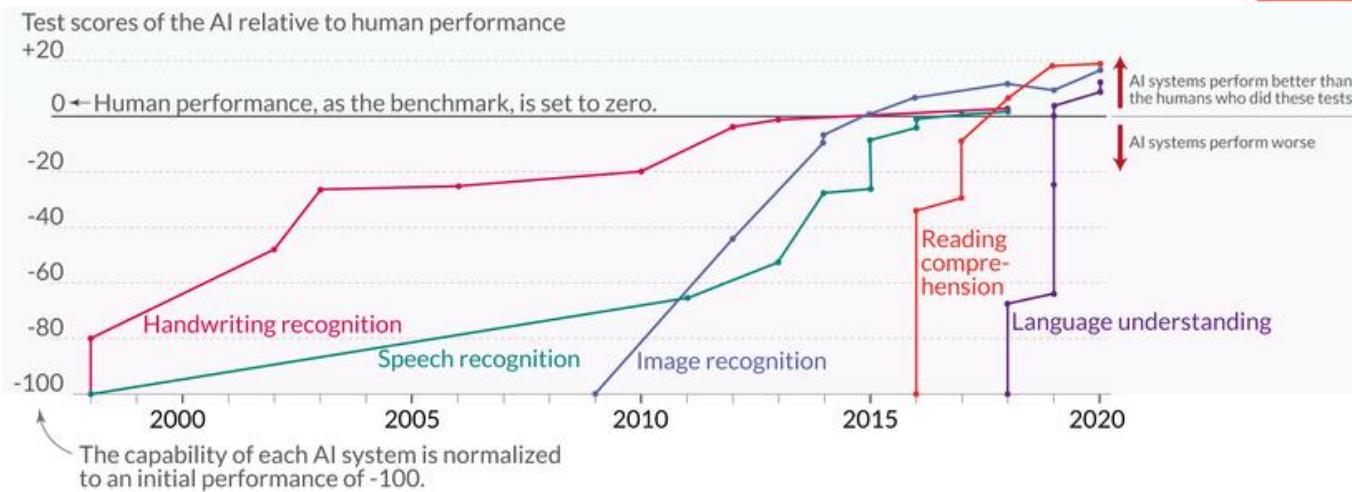


# Use cases.



# Introduction.

Language and image recognition capabilities of AI systems have improved rapidly



Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP  
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

Timeline of images generated by artificial intelligence  
These people don't exist. All images were generated by artificial intelligence.



OurWorldinData.org – Research and data to make progress against the world's largest problems. Licensed under CC-BY by the authors Charlie Giattino and Max Roser

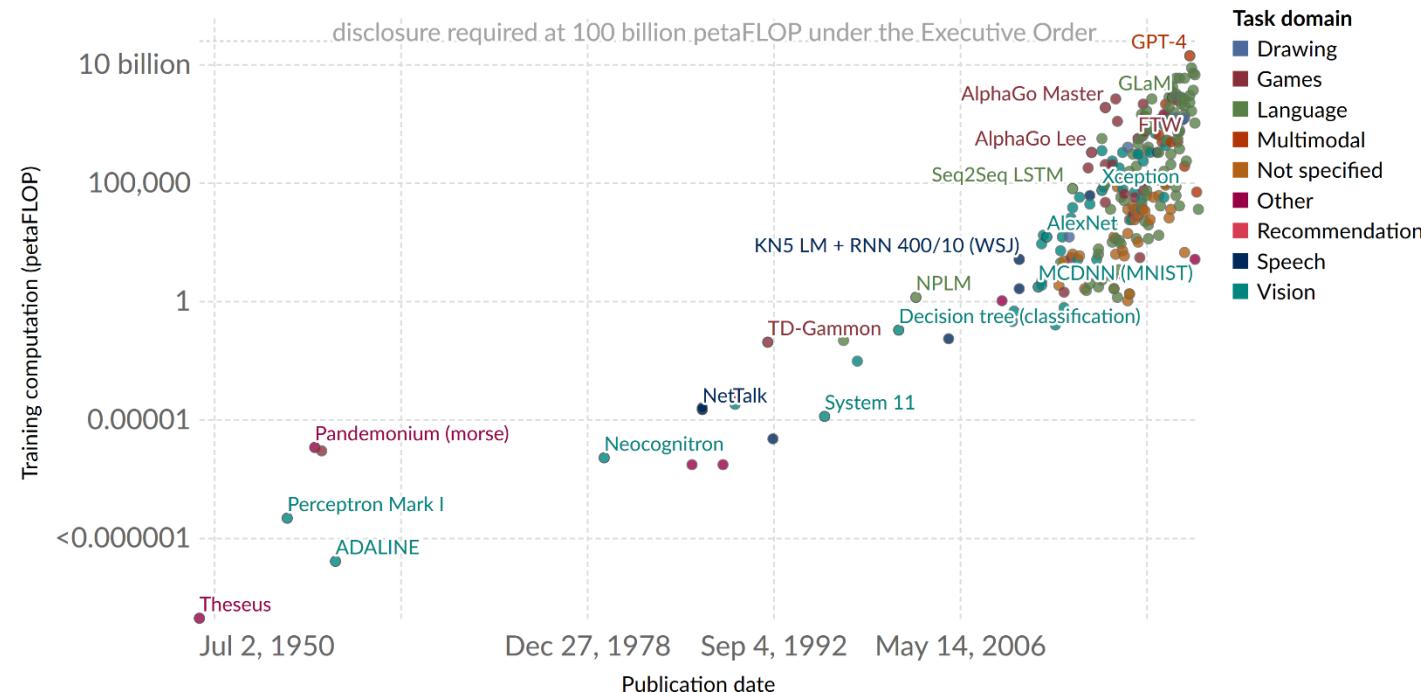


# Introduction.

## Computation used to train notable artificial intelligence systems

Our World  
in Data

Computation is measured in total petaFLOP, which is  $10^{15}$  floating-point operations<sup>1</sup> estimated from AI literature, albeit with some uncertainty. Estimates are expected to be accurate within a factor of 2, or a factor of 5 for recent undisclosed models like GPT-4.



Data source: Epoch (2023)

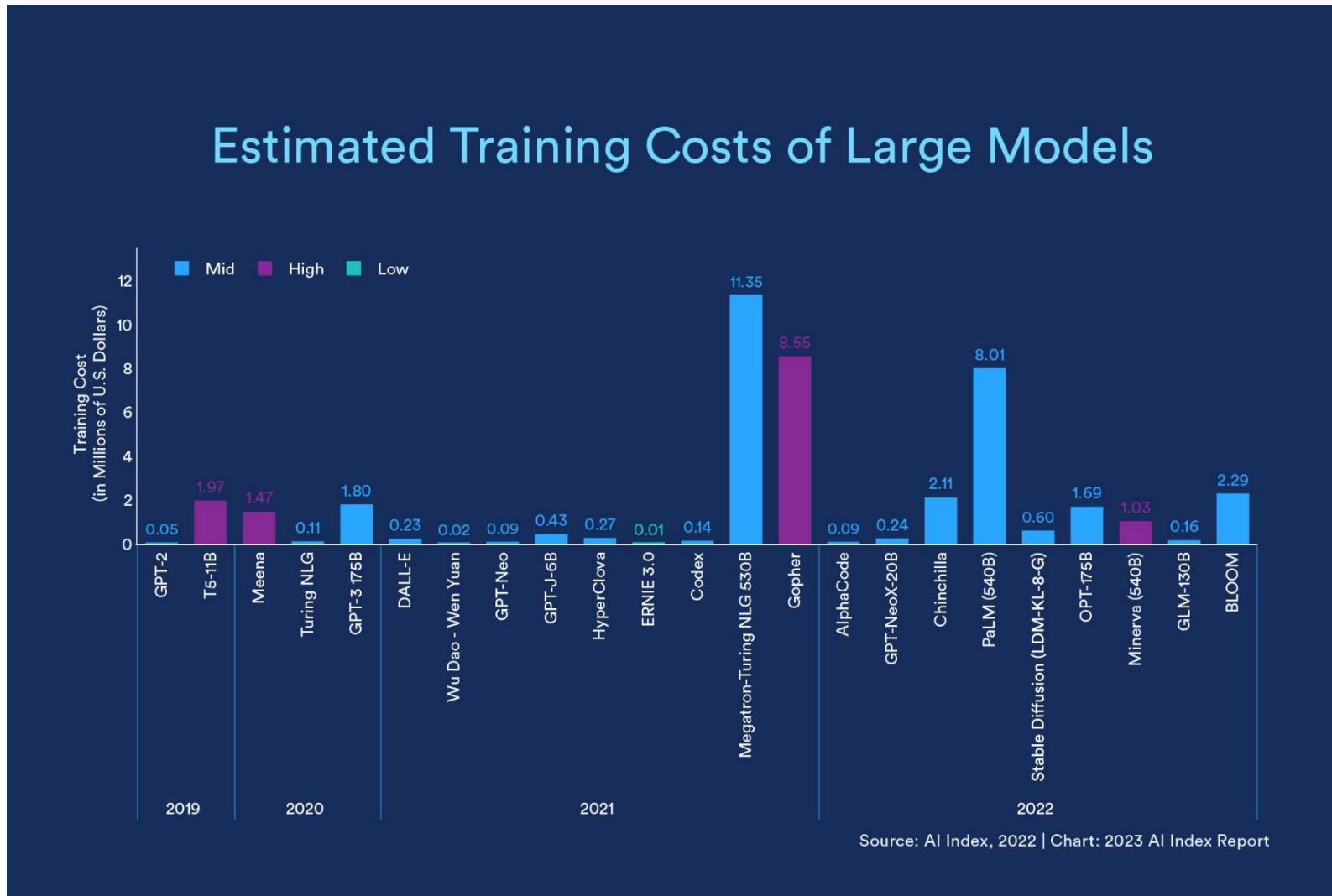
[OurWorldInData.org/artificial-intelligence](https://OurWorldInData.org/artificial-intelligence) | CC BY

Note: The Executive Order on AI refers to a directive issued by President Biden on October 30, 2023, aimed at establishing guidelines and standards for the responsible development and use of artificial intelligence within the United States.

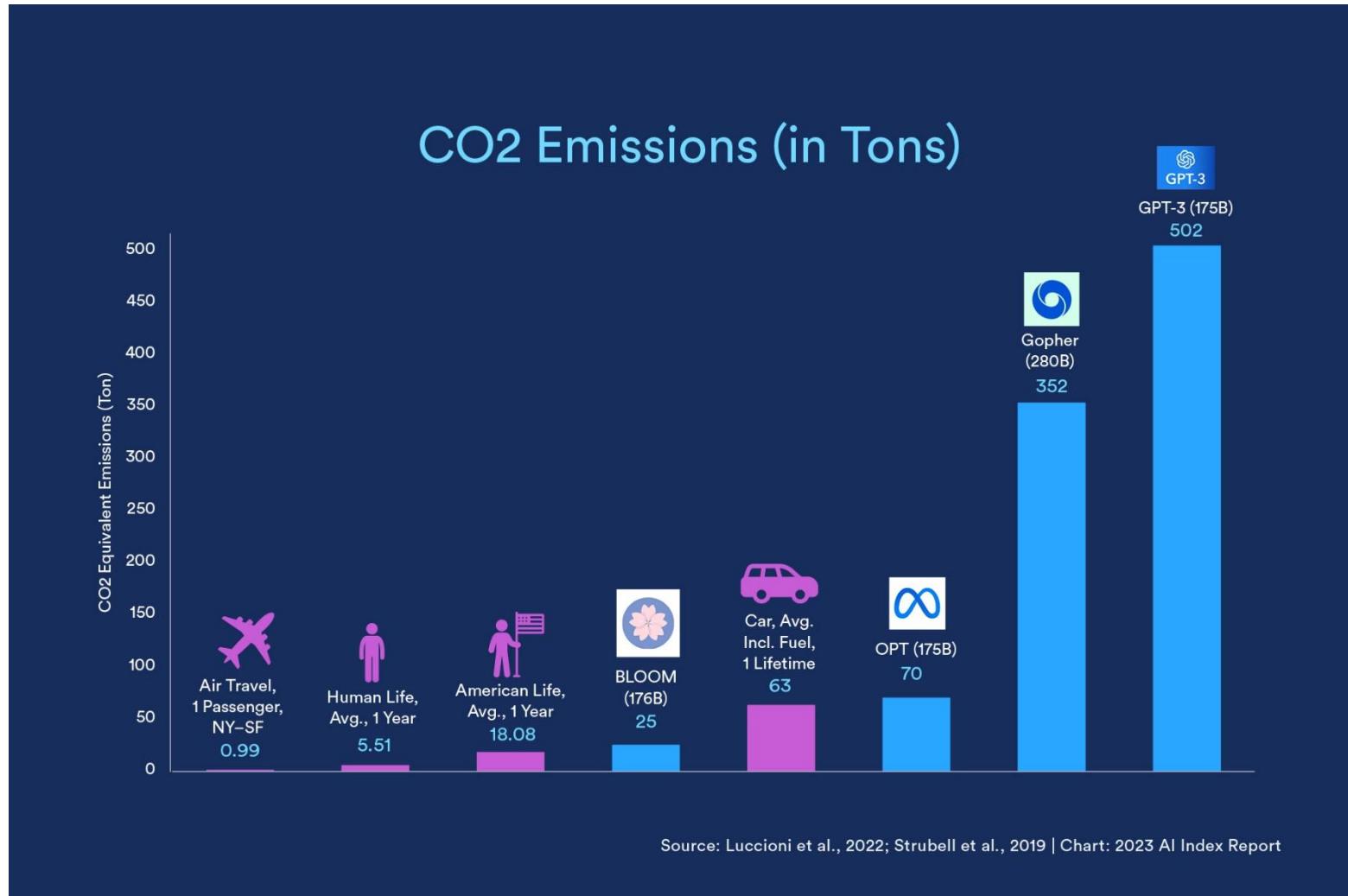
1. Floating-point operation: A floating-point operation (FLOP) is a type of computer operation. One FLOP is equivalent to one addition, subtraction, multiplication, or division of two decimal numbers.



# Introduction.



# Introduction.



# Failures.

## The final 11 seconds of a fatal Tesla Autopilot crash

A reconstruction of the wreck shows how human error and emerging technology can collide with deadly results



2.

# About this talk.

1. What to expect.
2. What NOT to expect.



# Introduction.

## AI for security

Refers to the application of artificial intelligence (AI) techniques and technologies to enhance and fortify cybersecurity measures.

- Intrusion detection
- Predictive analysis
- Malware detection
- Automated response systems
- ...

## Security of AI

Involves safeguarding artificial intelligence systems from potential vulnerabilities, attacks, and ethical considerations.

- Adversarial attacks
- Explainability and transparency
- Data privacy and confidentiality
- Ethics
- ...

# Introduction.

## What to expect

- Brief introduction to different attacks
- In-depth explanation of certain attacks
- State-of-the art methods
- Some demos!

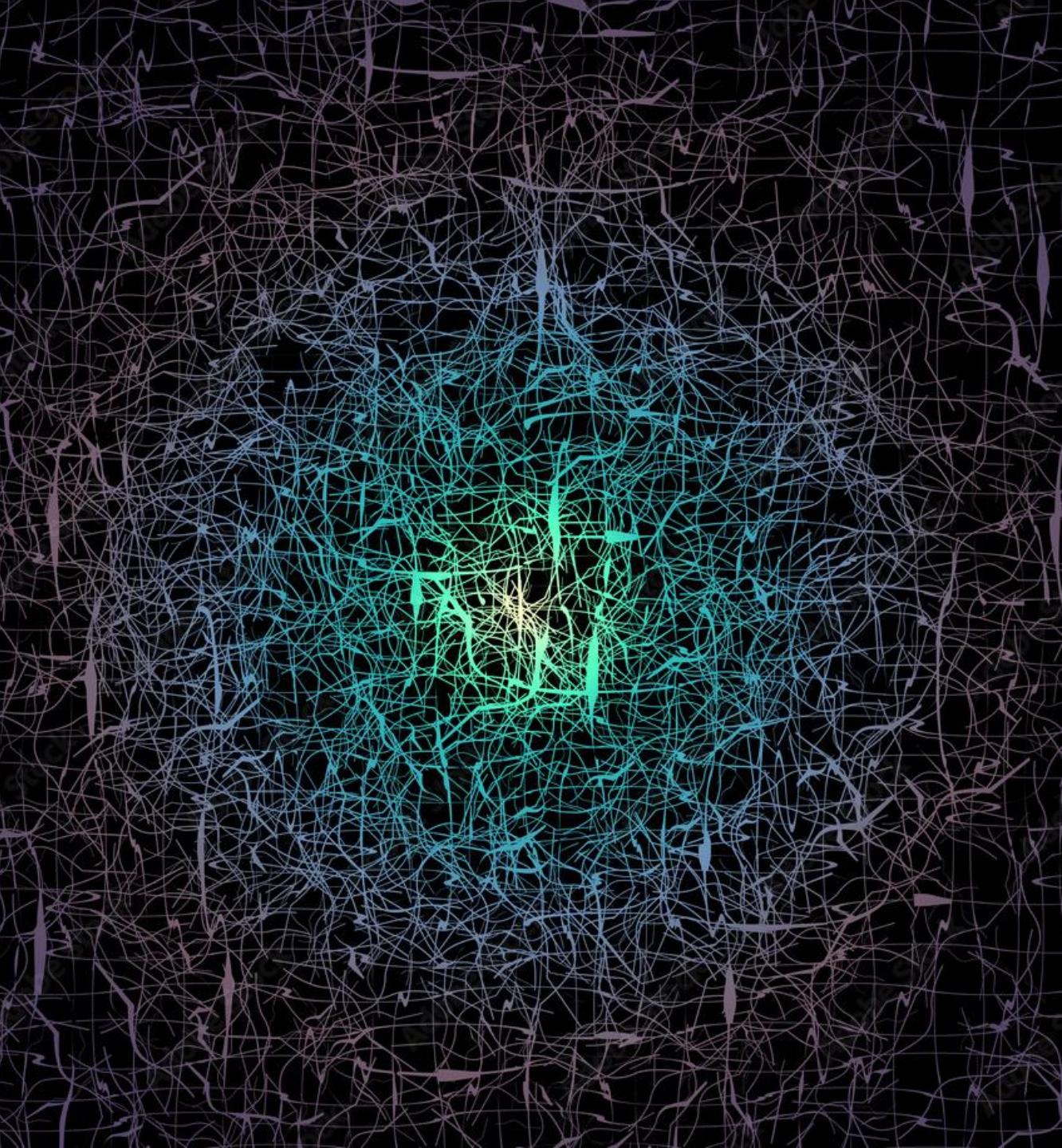
## What NOT to expect

- Hacking Chat-GPT
- Crashing a Tesla
- Lot of math (just some)
- Magic bullet solutions
- Apocalyptic scenarios

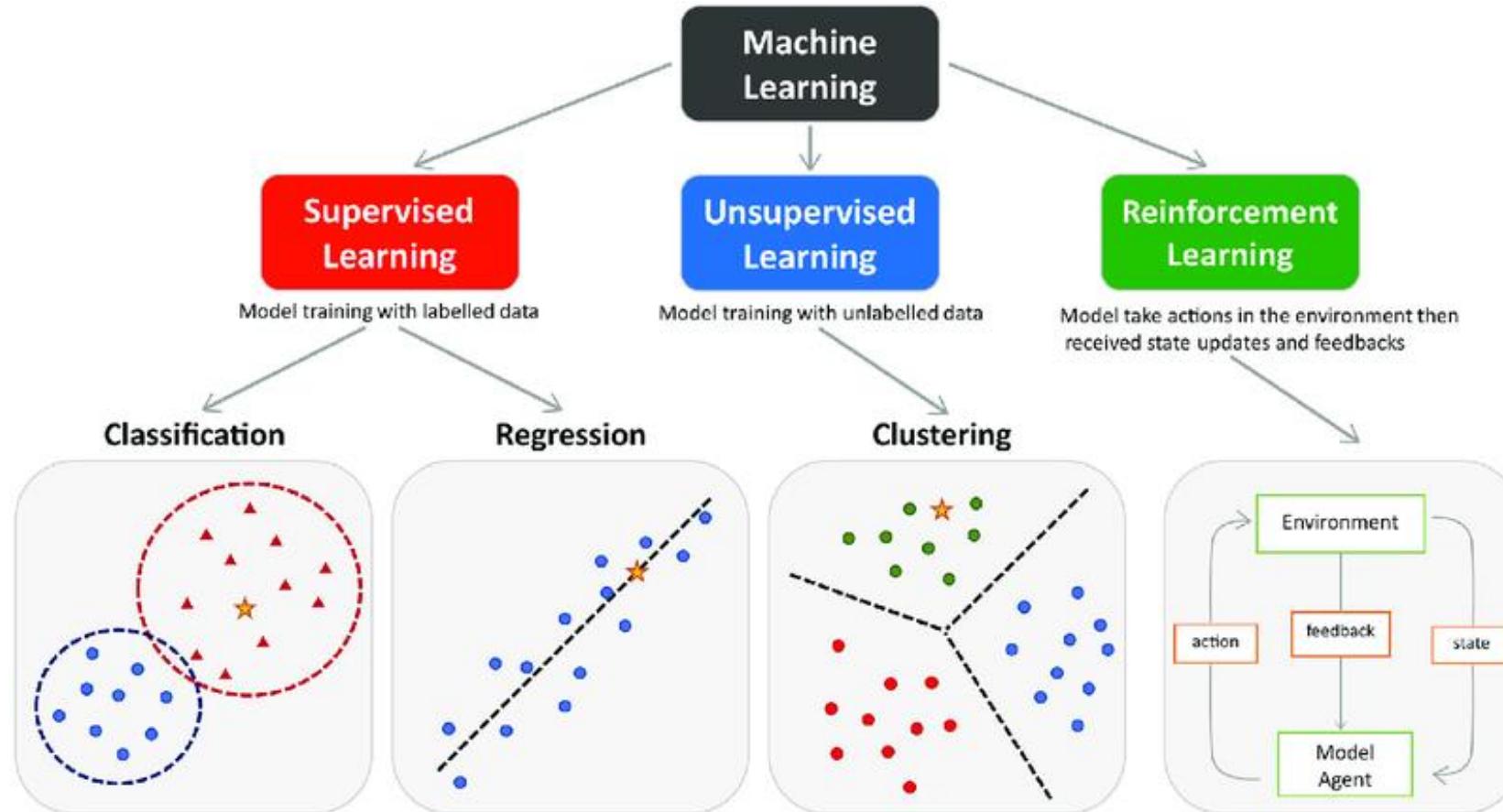
3.

# Deep Learning.

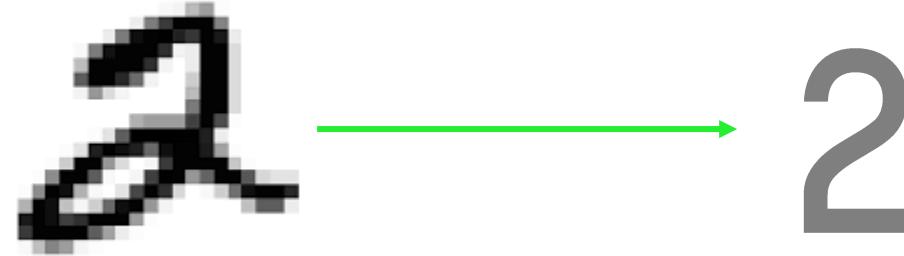
- 1. Introduction
- 2. How machines learn
  - 1. Gradients, Weights, and Inner Computations
  - 2. Gradient Descent



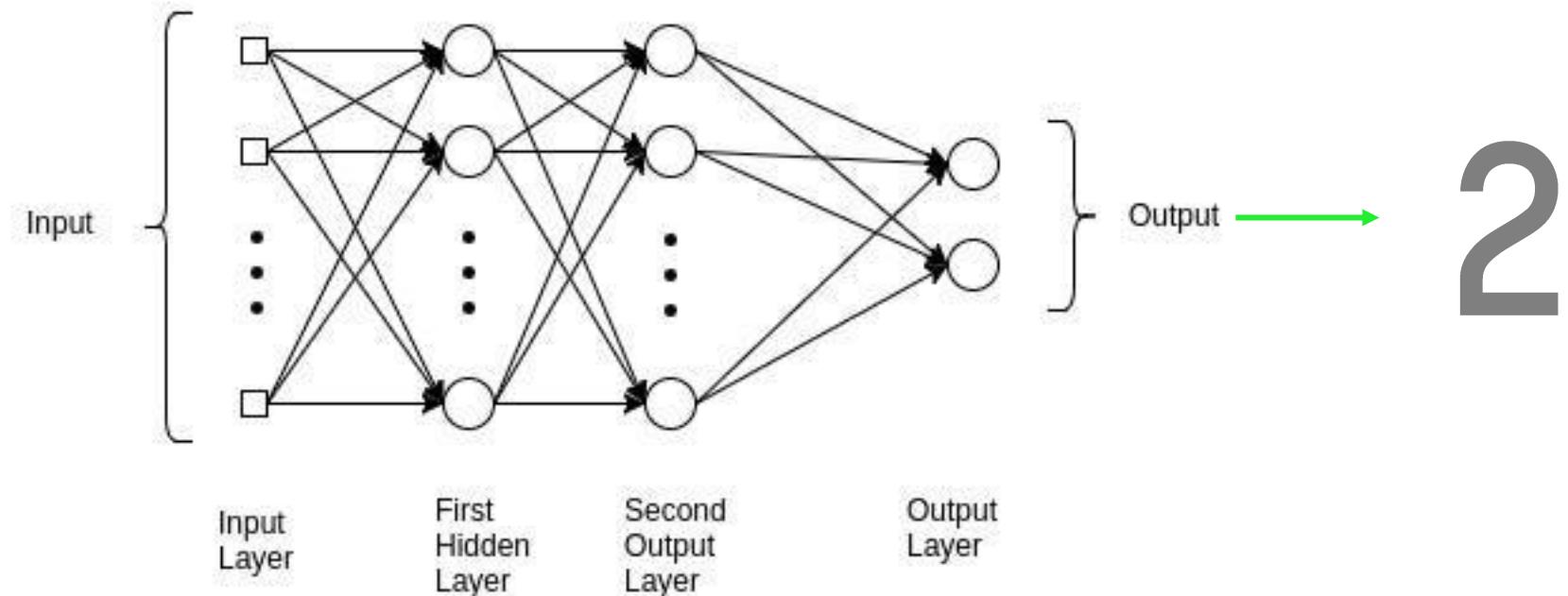
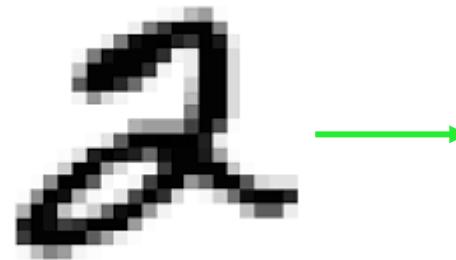
# Introduction.



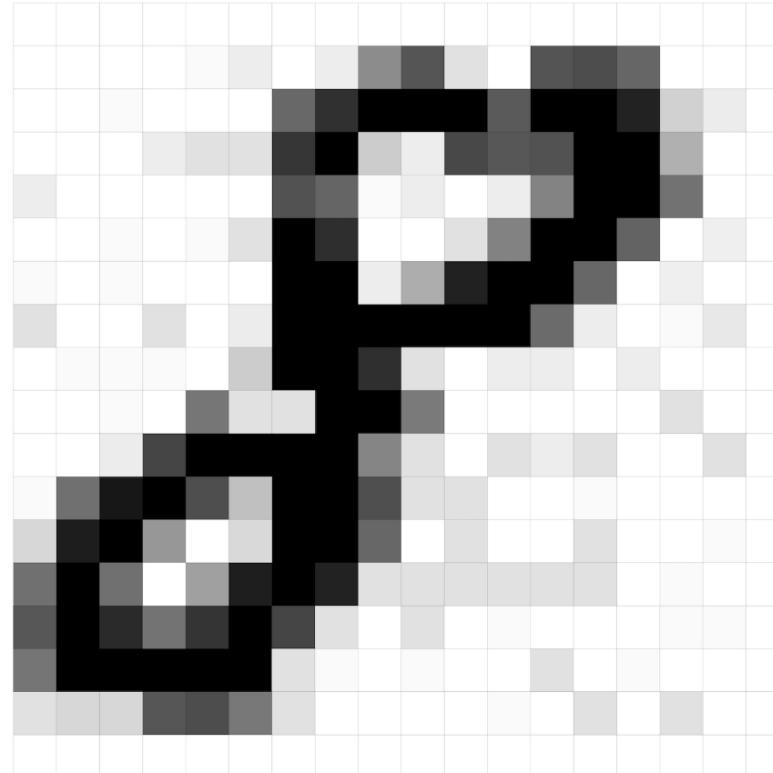
# Introduction.



# Introduction.

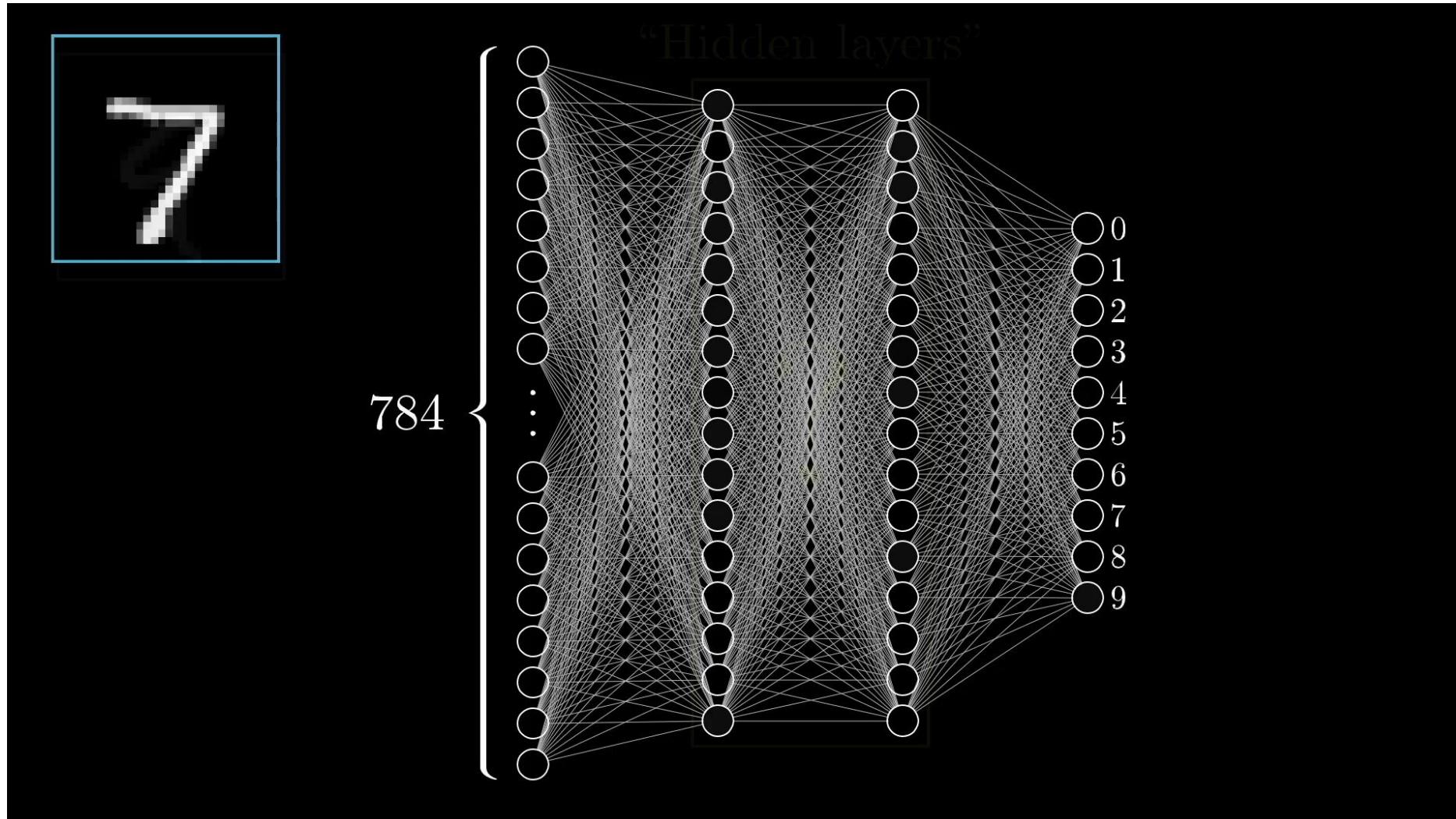


# Introduction.



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
0	0	1	0	0	0	41	160	252	256	230	160	254	236	203	11	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
0	0	0	16	9	9	148	250	45	21	184	159	154	255	233	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
10	0	0	0	0	0	143	147	3	10	0	10	122	250	254	106	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
0	0	3	0	3	10	236	216	0	0	38	109	247	240	169	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
1	0	2	0	0	0	252	253	23	62	224	241	255	164	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
6	0	0	4	0	8	254	250	250	228	254	234	112	28	0	2	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
0	1	1	4	0	21	254	250	126	6	0	10	14	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
0	0	4	0	163	8	8	250	229	120	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
3	79	240	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
125	255	141	0	87	244	255	208	8	8	8	8	8	8	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	23	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

# Introduction.



# How machines learn.

- Training sets the **parameters** of the neural network (NN).
- An **optimal** set of parameters makes the NN work great.
- We need **data**.
- Lots of data.
- In **supervised learning**:
  - Data is labelled (classes).
  - We call this a **Dataset**.



# How machines learn.

- Train on the training set.
- Evaluate on a holdout test set.
- Evaluating measures how good the model is doing. (Generalization)
- Metrics:
  - Accuracy
  - ROC curve
  - ...

Training set

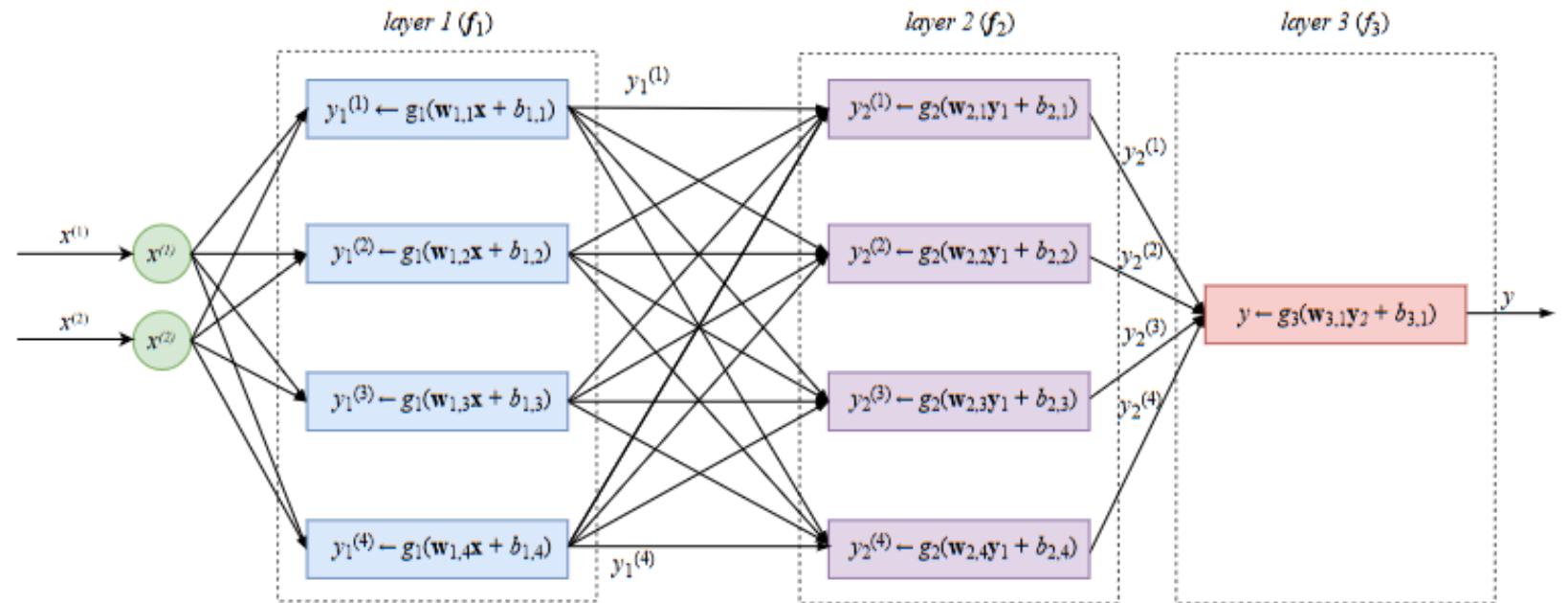
0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

Test set

0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9

# How machines learn.

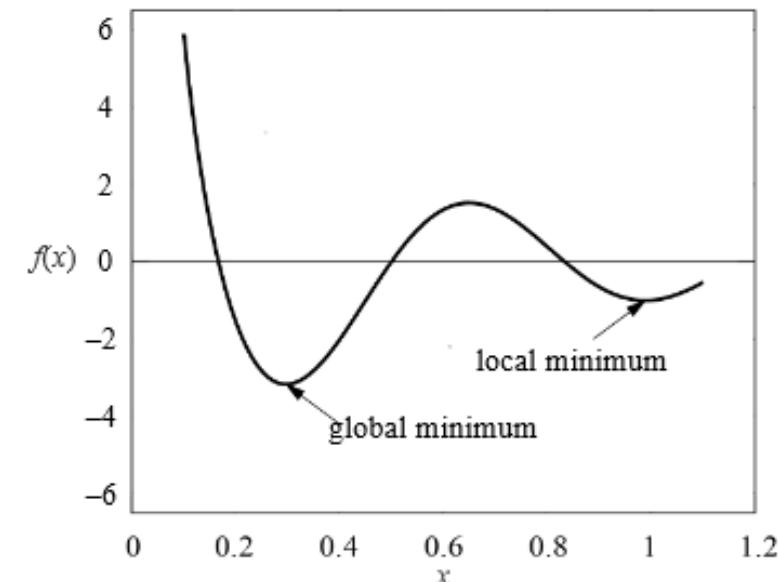
- Input ( $\mathbf{x}$ )
- Layers
  - Input
  - Hidden
  - Output
- Neurons\*
  - Weights ( $\mathbf{w}$ )
- Activation functions ( $g(\cdot)$ )
- Output ( $y$ )



\* Each connection between neurons has a weight, rather than each neuron.

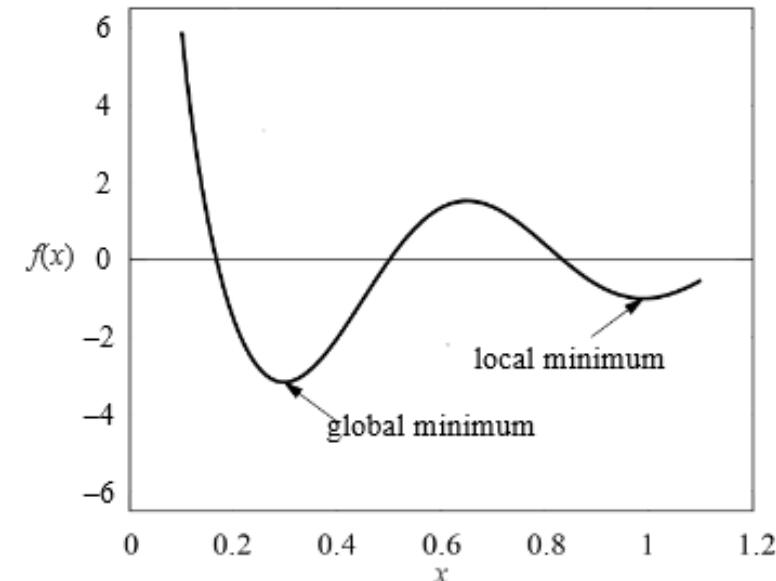
# How machines learn.

- Training is pretty much about finding the **minimum** of a function.
- Derivatives:
  - The derivative  $f'$  of a function  $f$  describes how fast  $f$  grows or decreases.
  - Chain rule.
  - In DL we use **partial derivatives**, since we have  $n$  dimensions.
  - To the vector of partial derivatives we name it **Gradients** ( $\nabla$ ).

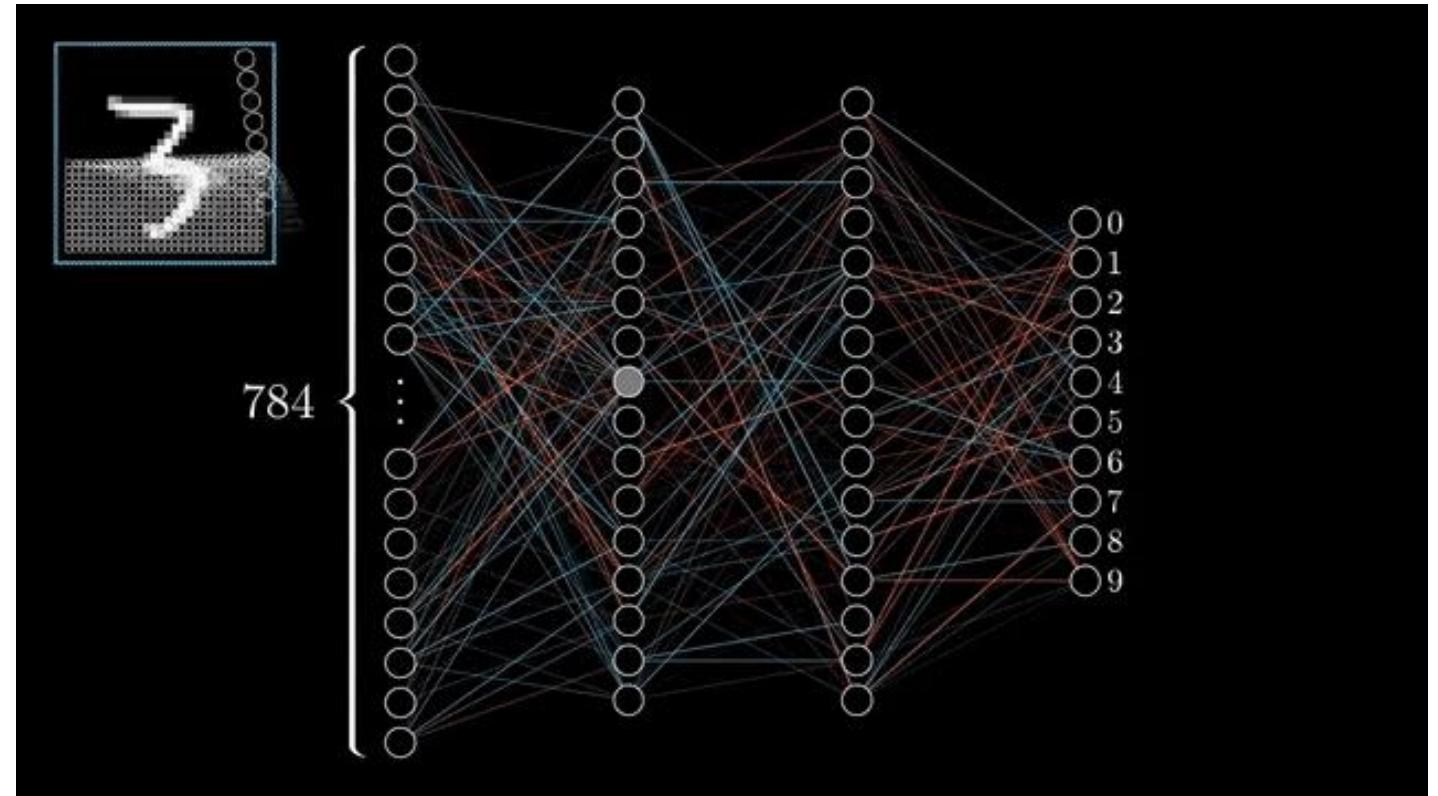
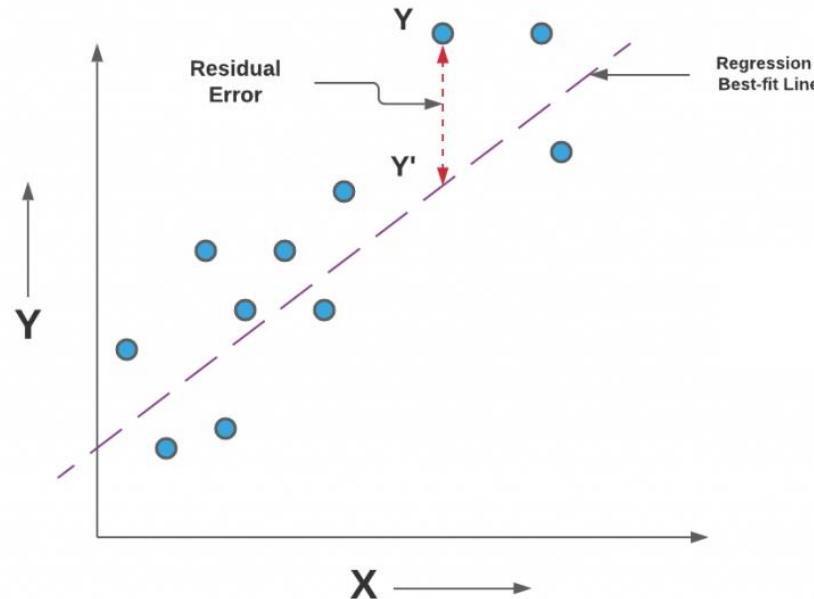


# How machines learn.

- Training is about finding the **optimal parameters** (weights).
- The optimal parameters are found by finding the **minimum** of a function (minimum cost).
- The **gradients** point towards the steepest ascent.
- The **minimum** is found using the gradients.

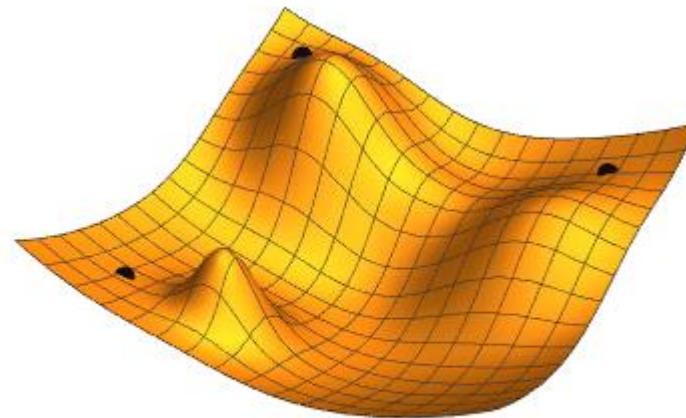


# How machines learn.

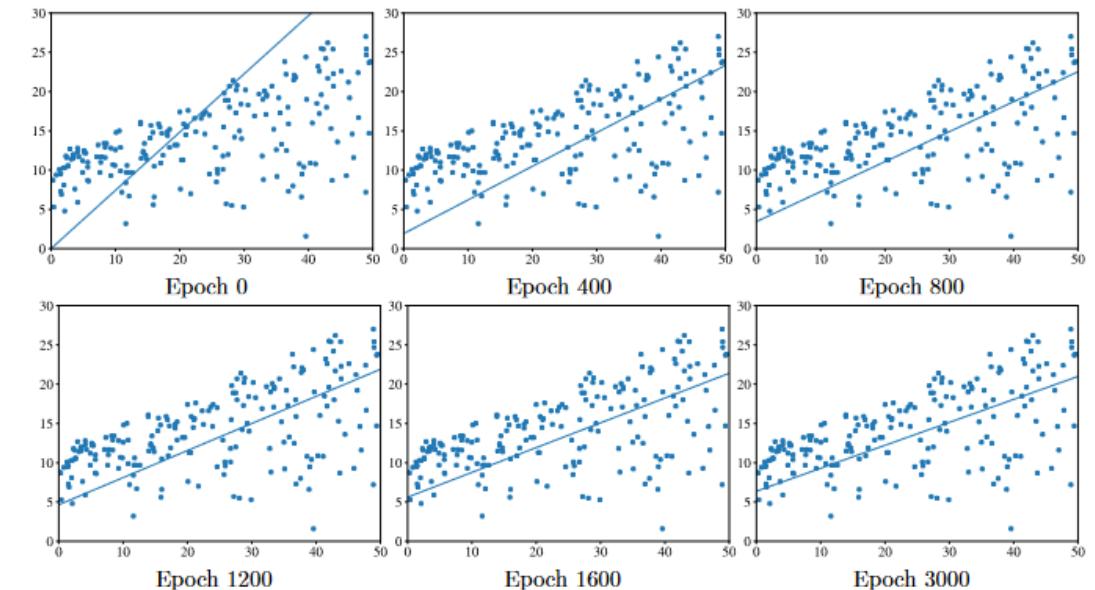


# How machines learn.

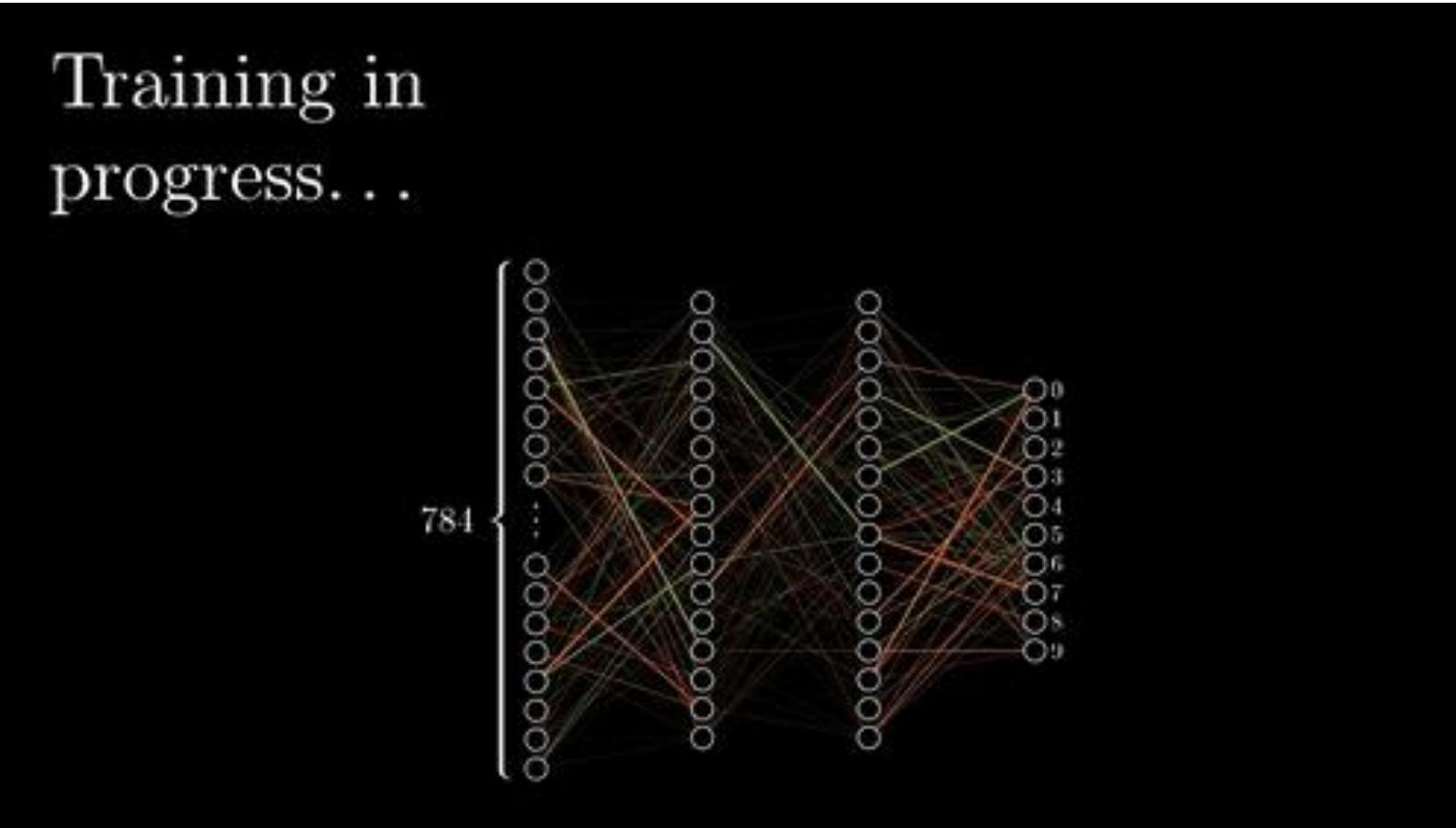
Gradient descent



Optimizing the weights (training)



# How machines learn.



4.

# Privacy in Deep Learning.

1. Introduction
2. Inference Attacks
3. Model Inversion
4. Model Extraction

# What's privacy?.

*"Data privacy is a discipline intended to keep data safe against improper access, theft or loss".*

Attacks to privacy try to extract information from the model, e.g., recover the data used during training.

LONG LIVE THE REVOLUTION.  
OUR NEXT MEETING WILL BE  
AT THE DOCKS AT MIDNIGHT  
ON JUNE 28 TAB

AHA, FOUND THEM!



WHEN YOU TRAIN PREDICTIVE MODELS  
ON INPUT FROM YOUR USERS, IT CAN  
LEAK INFORMATION IN UNEXPECTED WAYS.

# Types of attacks.



## Model stealing (model extraction) [1]

Model extraction attacks target the confidentiality of a victim model (architecture and its parameters) deployed on a remote service.



## Membership inference [3]

Given a data point, the adversary infers whether this data point is in the training dataset of the target model by querying it.



## Model inversion [2]

Given a trained model, the attacker aims to partially or entirely reconstruct the training data.

Original



Extracted

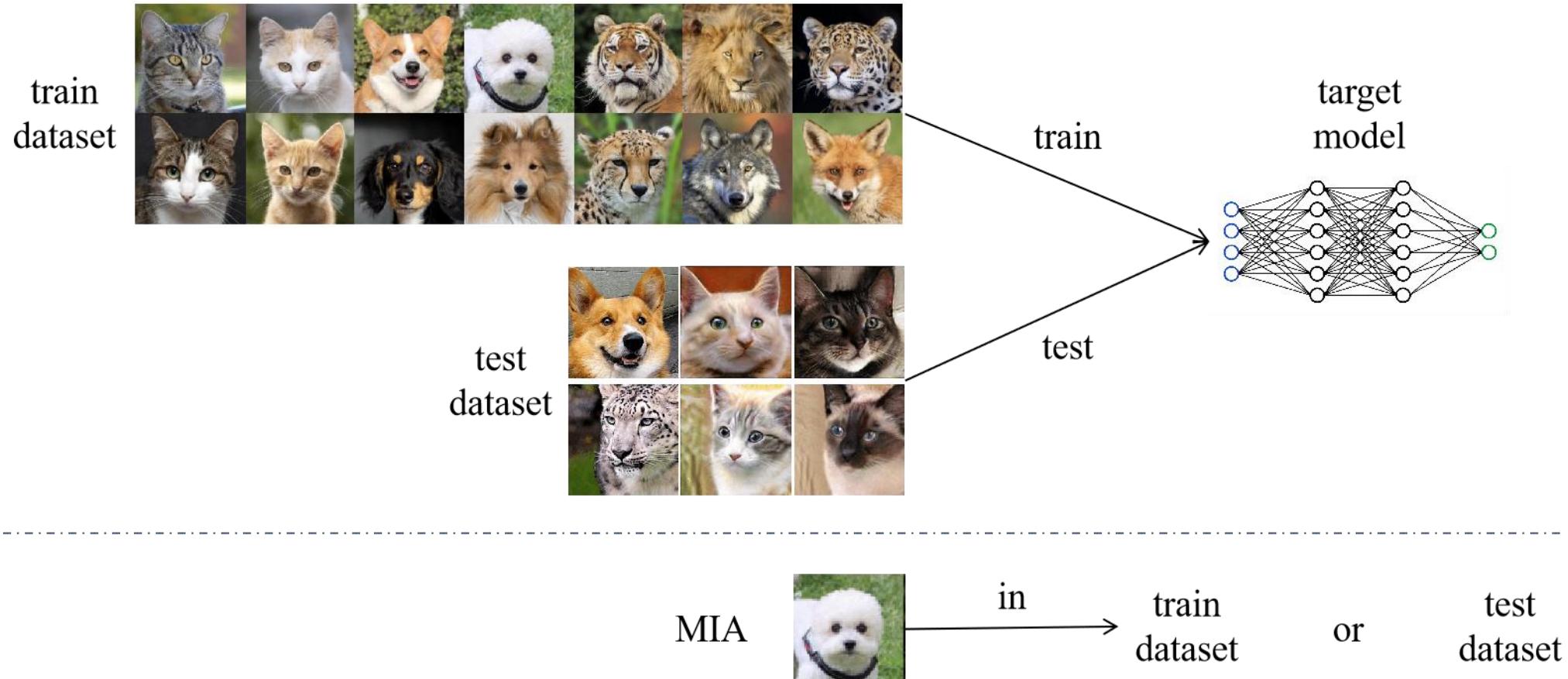


[1] Jagielski et al. "High accuracy and high-fidelity extraction of neural networks." USENIX Security 2020.

[2] Fredrikson, Matt, Somesh Jha, and Thomas Ristenpart. "Model inversion attacks that exploit confidence information and basic countermeasures." *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*. 2015.

[3] Shokri, Reza, et al. "Membership inference attacks against machine learning models." *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017.

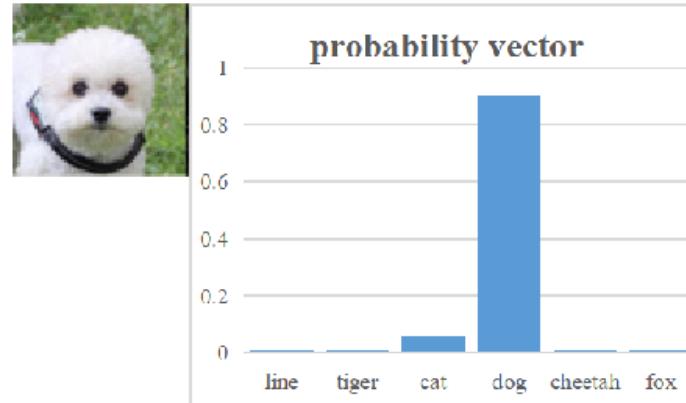
# Membership inference.



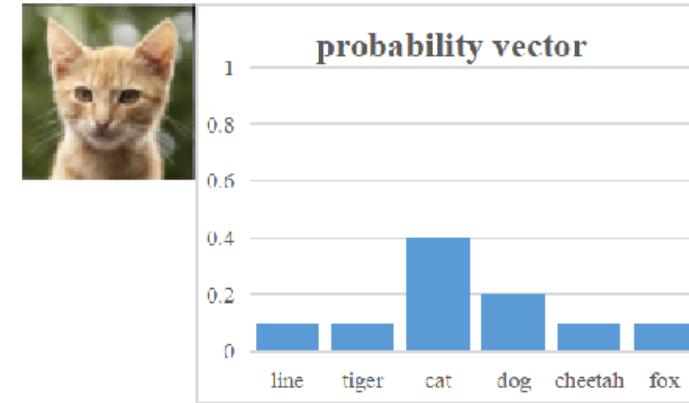
# Membership inference.

## Any solution?

train data point



test data point



# Types of attacks.



## Model stealing (model extraction) [1]

Model extraction attacks target the confidentiality of a victim model (architecture and its parameters) deployed on a remote service.



## Model inversion [2]

Given a trained model, the attacker aims to partially or entirely reconstruct the training data.

Original



Extracted



[1] Jagielski et al. "High accuracy and high-fidelity extraction of neural networks." USENIX Security 2020.



## Membership inference [3]

Given a data point, the adversary infers whether this data point is in the training dataset of the target model by querying it.



## Any defenses?

Adding noise (differential privacy).

Encryption.

Output filtering.

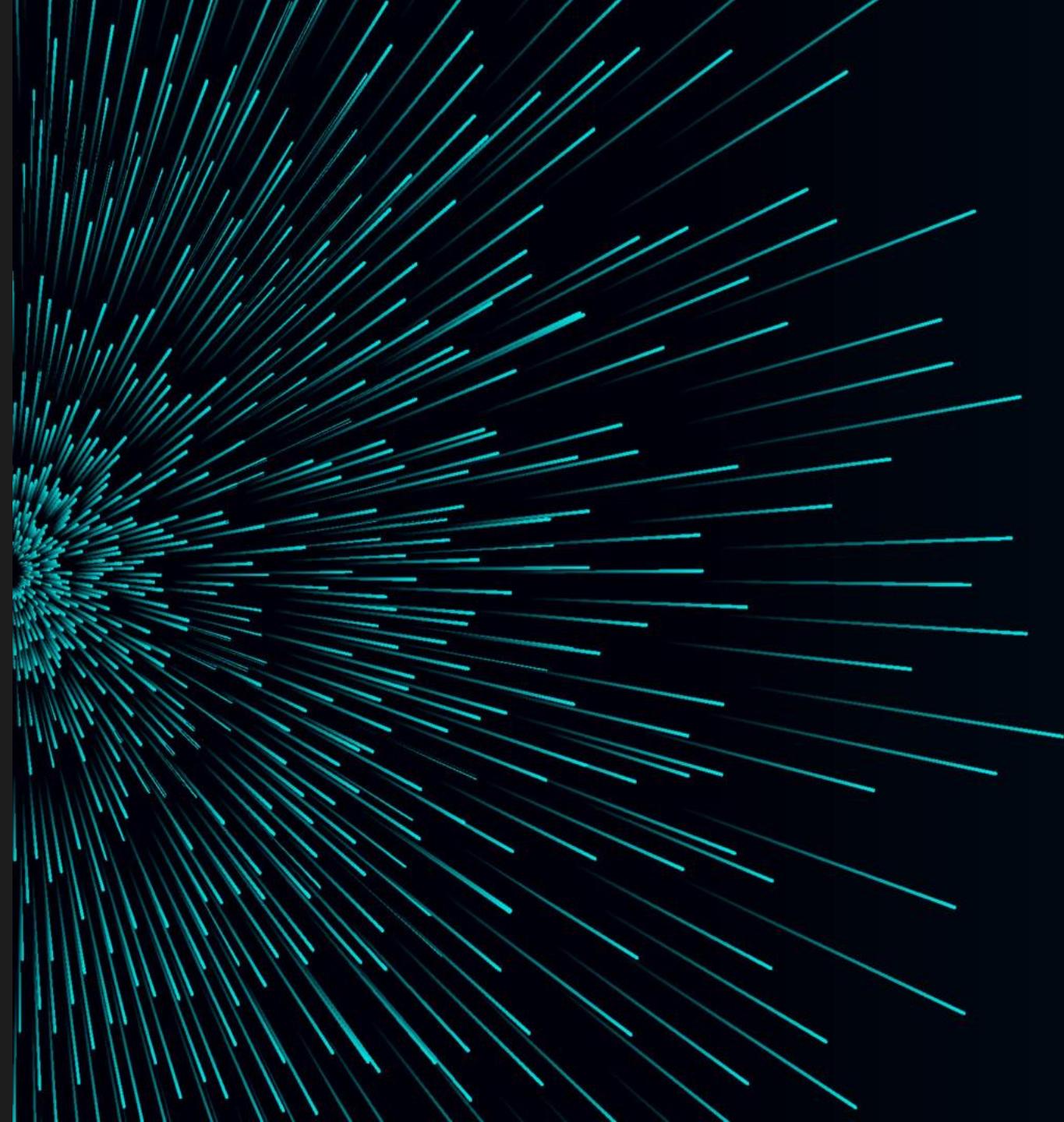
...

[3] Shokri, Reza, et al. "Membership inference attacks against machine learning models." *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017.

5.

# Security in Deep Learning.

1. Adversarial examples
2. Backdoor attacks



# What's security?.

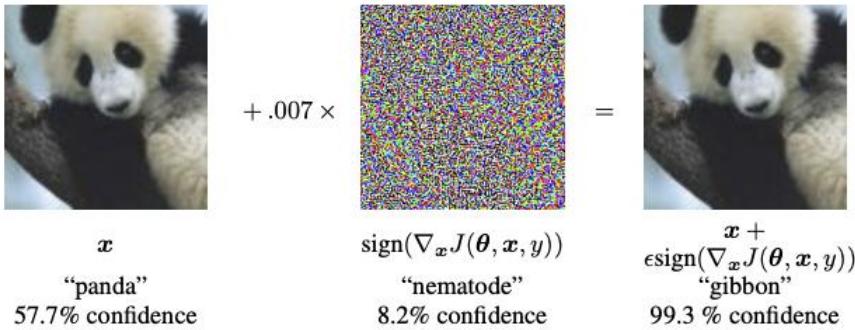
“Security is a comprehensive discipline aimed at safeguarding assets, including data, information systems, and physical resources, against unauthorized access, damage, disruption, or theft”.

Attacks on the ML models' security try to make the model misbehave, e.g., denial of service or targeted misclassification.



# Type of attacks.

## Adversarial examples



Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples.” *arXiv preprint arXiv:1412.6572* (2014).



S. Thys, W. V. Ranst, and T. Goedemé, “Fooling automated surveillance cameras: Adversarial patches to attack person detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2019

## Backdoor attacks



Doan, Khoa, et al. “Lira: Learnable, imperceptible and robust backdoor attacks.” *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.

# Type of attacks.

## Adversarial examples

- **Objective:**
  - Generate inputs that are intentionally crafted to mislead the model during inference.
  - Goal is to cause misclassification or a wrong prediction without modifying the model's parameters.
- **Method:**
  - Perturbations are added to input data, often imperceptible to humans.
  - Adversarial attacks are typically focused on exploiting weaknesses in the model's decision boundary.

## Backdoor attacks

- **Objective:**
  - Introduce a specific trigger pattern during the training phase that, when present in the input during inference, causes the model to behave maliciously.
  - Goal is to have a model exhibit unwanted behavior when presented with a specific, often rare, input pattern.
- **Method:**
  - A small, carefully chosen subset of the training data is manipulated to include the trigger pattern.
  - The model *learns* to associate this trigger pattern with a specific malicious outcome.

# Type of attacks.

## Adversarial examples

### Knowledge

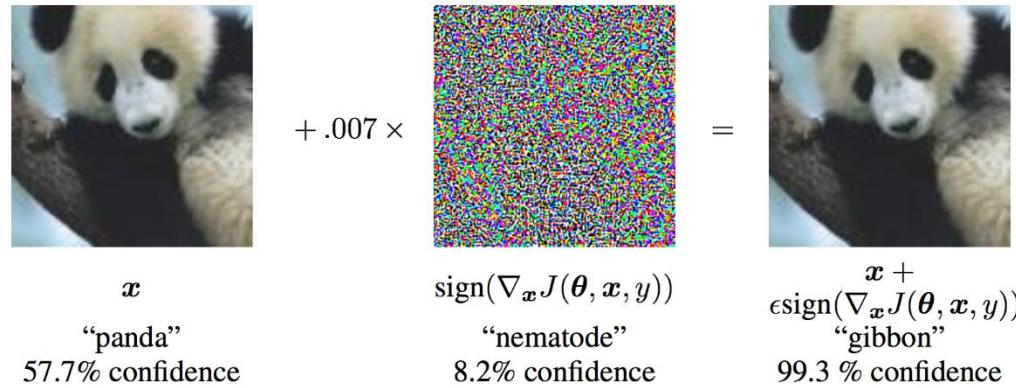
- A *white-box* attack assumes the attacker has full knowledge and access to the model, including architecture, inputs, outputs, and weights.
- A *black-box* attack assumes the attacker only has access to the inputs and outputs of the model, and knows nothing about the underlying architecture or weights.

### Goal

- A goal of *misclassification* means the adversary only wants the output classification to be wrong but does not care what the new classification is.
- A *source/target misclassification* means the adversary wants to alter an image that is originally of a specific source class so that it is classified as a specific target class.

# Type of attacks.

## Fast Gradient Sign Method (FGSM) [1]



The attack is remarkably powerful, and yet intuitive. It is designed to attack neural networks by leveraging the way they learn, *gradients*.

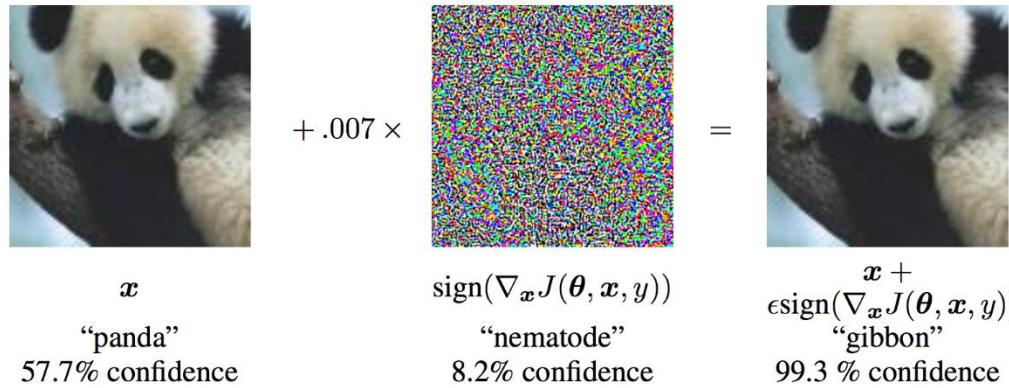
The idea is simple, rather than working to minimize the loss by adjusting the weights based on the backpropagated gradients, the attack *adjusts the input data to maximize the loss* based on the same backpropagated gradients.

In other words, the attack uses the gradient of the loss w.r.t the input data, then adjusts the input data to maximize the loss.

[1] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." *arXiv preprint arXiv:1412.6572* (2014).

# Type of attacks.

## Fast Gradient Sign Method (FGSM) [1]



- $\boldsymbol{x}$  Original input
- $\boldsymbol{x}'$  Perturbed image
- $\boldsymbol{y}$  Ground truth label
- $\boldsymbol{\theta}$  Model parameters
- $J$  Loss function
- $\nabla_{\boldsymbol{x}}$  Gradients
- sign** The sign (+,-) of the gradients.
- $\epsilon$  Noise step

The input data is adjusted by a small step (0.007 in the picture) in the direction (i.e.  $\text{sign}(\nabla_x J(\theta, x, y))$ ) that will maximize the loss.

The resulting perturbed image,  $\boldsymbol{x}'$ , is then *misclassified* by the target network as a “*gibbon*” when it is still clearly a “*panda*”.

[1] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." *arXiv preprint arXiv:1412.6572* (2014).

# Type of attacks.

## Fast Gradient Sign Method (FGSM) [1]

$x$   
“panda”  
57.7% confidence

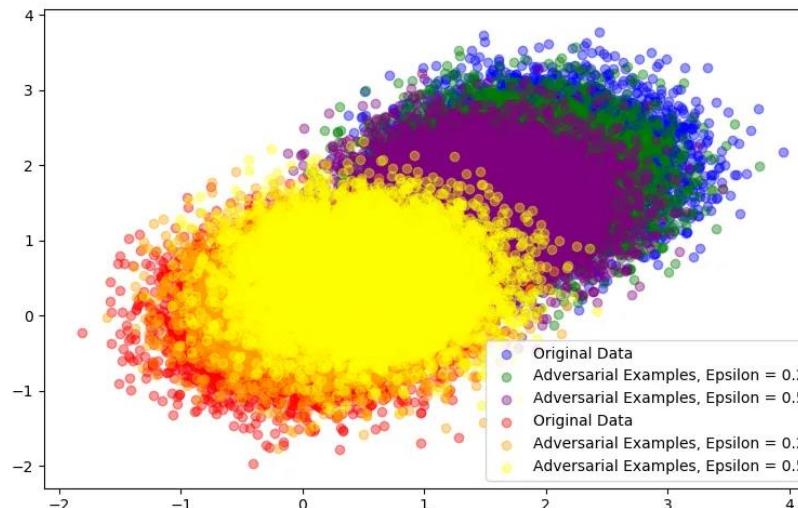
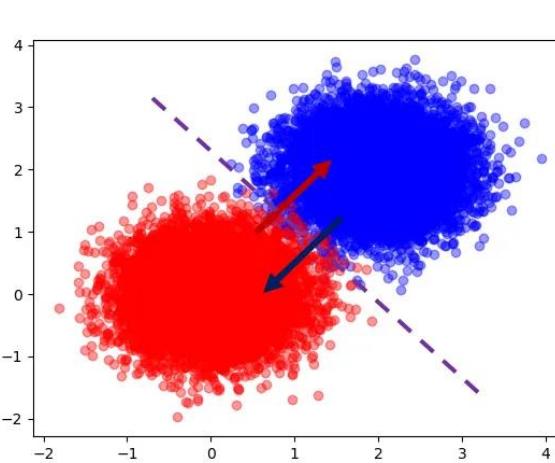
$+ .007 \times$

$\text{sign}(\nabla_{\mathbf{x}} J(\theta, \mathbf{x}, y))$   
“nematode”  
8.2% confidence

$=$

$\mathbf{x} + \epsilon \text{sign}(\nabla_{\mathbf{x}} J(\theta, \mathbf{x}, y))$   
“gibbon”  
99.3 % confidence

$\mathbf{x}$  Original input  
 $\mathbf{x}'$  Perturbed image  
 $\mathbf{y}$  Ground truth label  
 $\theta$  Model parameters  
 $J$  Loss function  
 $\nabla_{\mathbf{x}}$  Gradients  
**sign** The sign (+,-) of the gradients.  
 $\epsilon$  Noise step

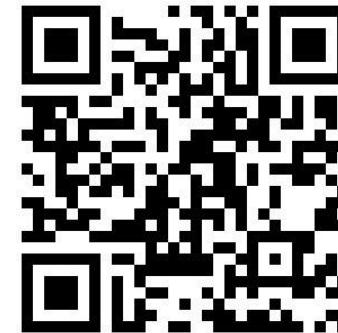


[1] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." *arXiv preprint arXiv:1412.6572* (2014).

# Type of attacks.

Fast Gradient Sign Method (FGSM)

## DEMO



<https://t.ly/Yj2-4>

# Type of attacks.

## Backdoor attacks

### Knowledge

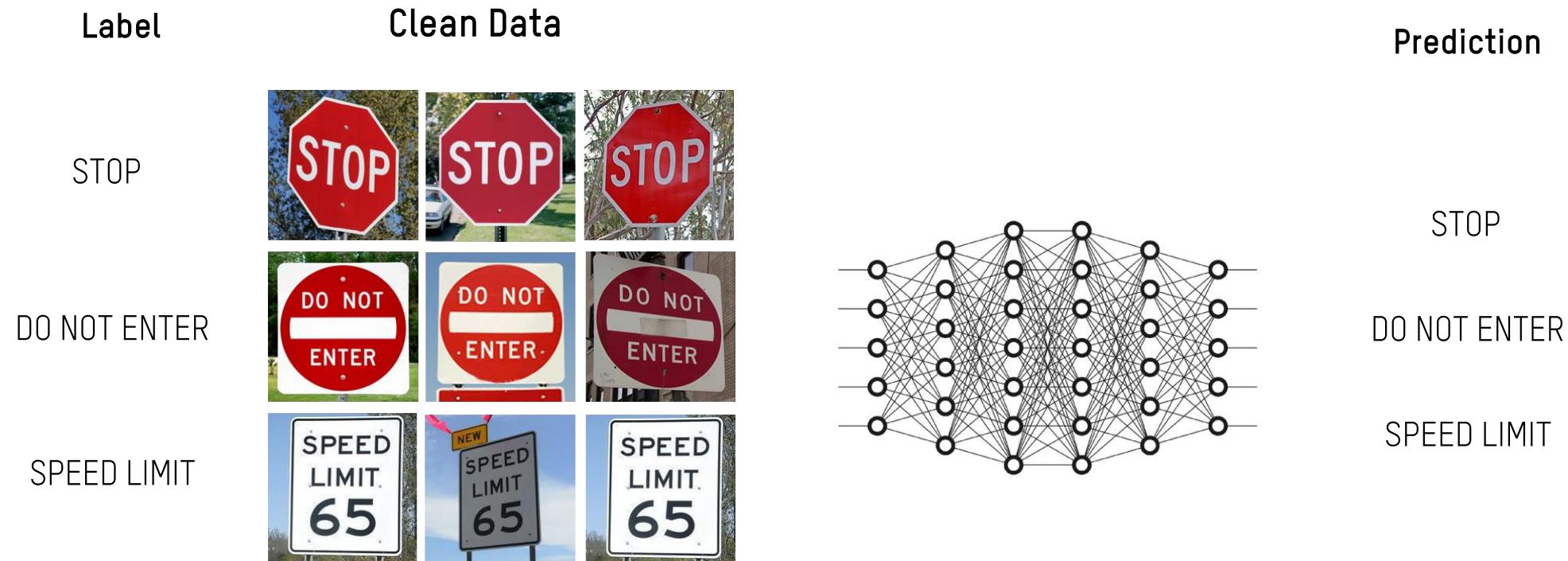
- Backdoor attacks are *training time* attacks.
- The attacker usually has more knowledge than in adversarial examples.
- The attacker may have access to the dataset.

### Goal

- A goal of *misclassification* means the adversary only wants the output classification to be wrong but does not care what the new classification is.
- A *source/target misclassification* means the adversary wants to alter an image that is originally of a specific source class so that it is classified as a specific target class.

# Type of attacks.

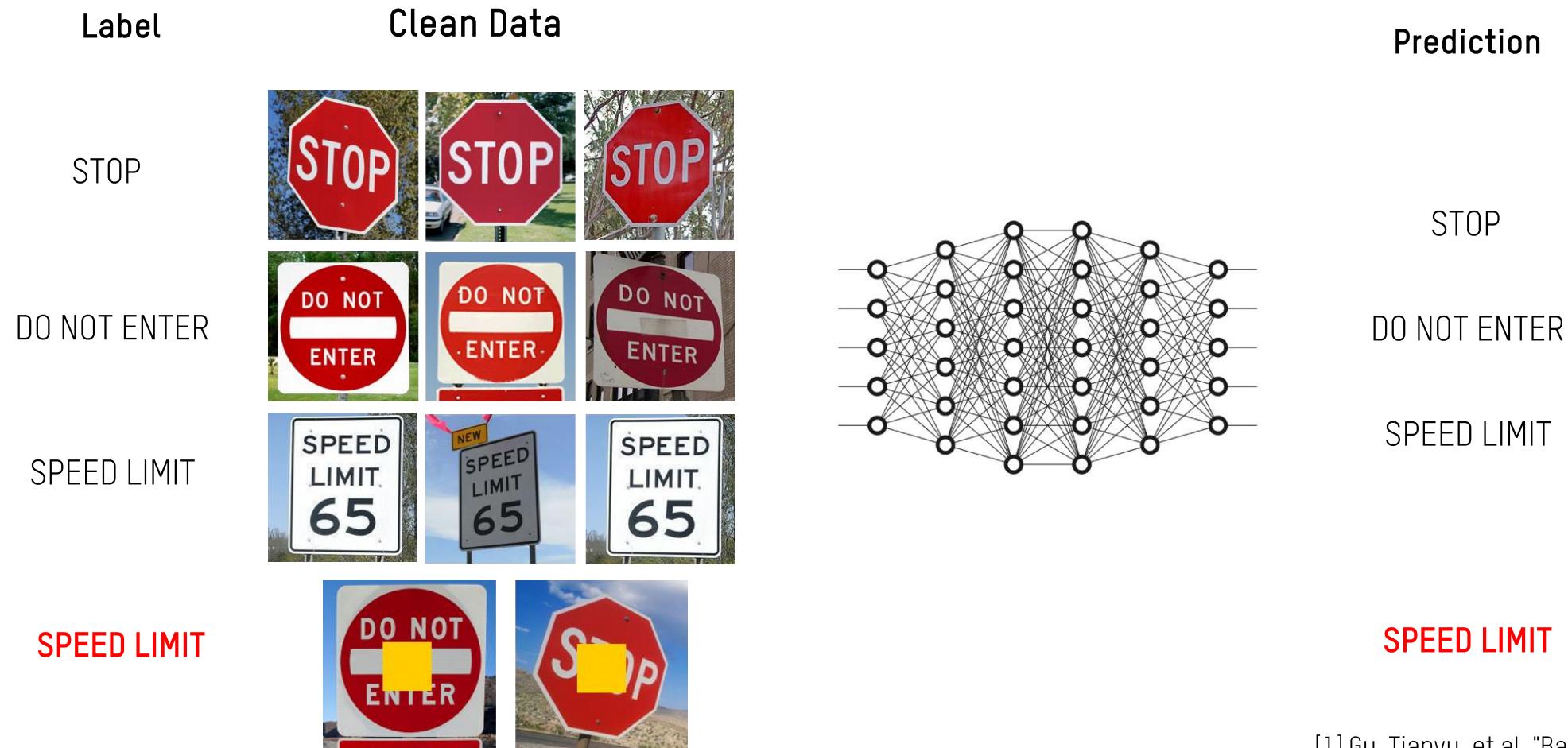
## Backdoor attacks [1]



[1] Gu, Tianyu, et al. "Badnets: Evaluating backdooring attacks on deep neural networks." *IEEE Access* 7 (2019): 47230-47244.

# Type of attacks.

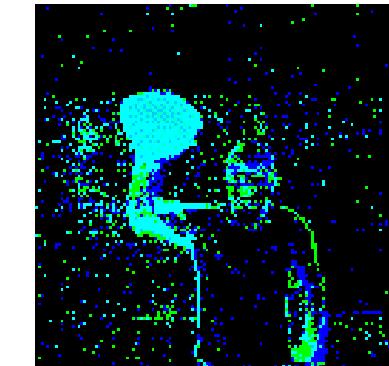
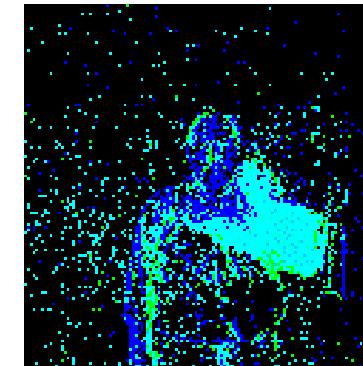
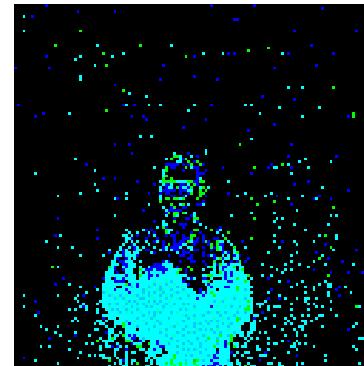
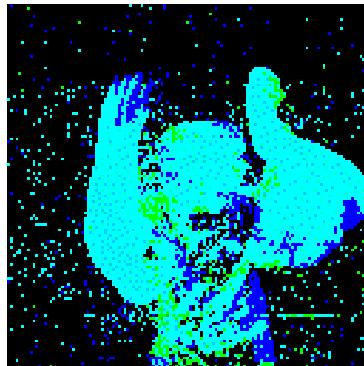
## Backdoor attacks [1]



[1] Gu, Tianyu, et al. "Badnets: Evaluating backdooring attacks on deep neural networks." *IEEE Access* 7 (2019): 47230-47244.

# Type of attacks.

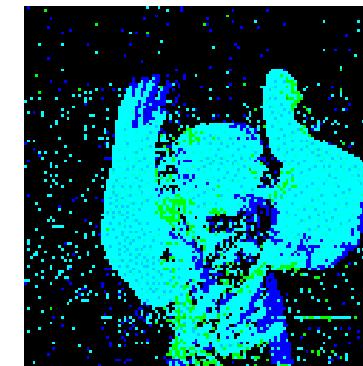
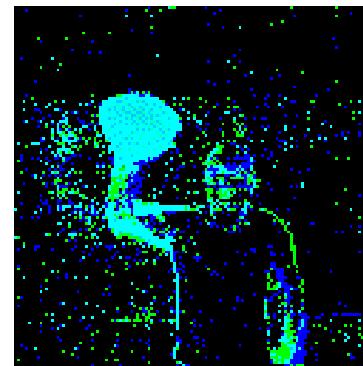
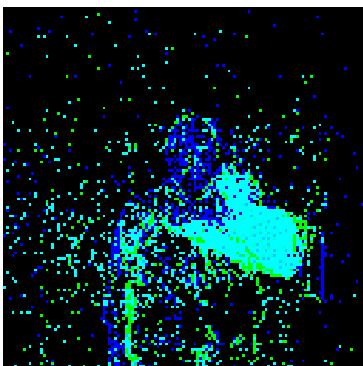
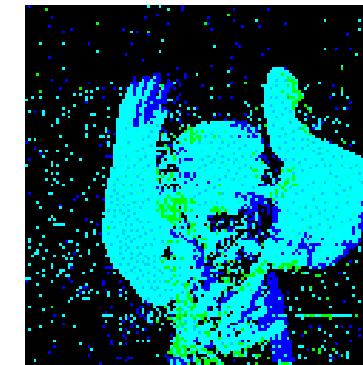
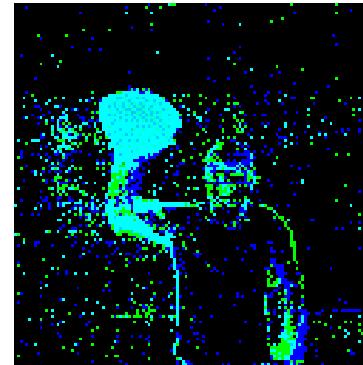
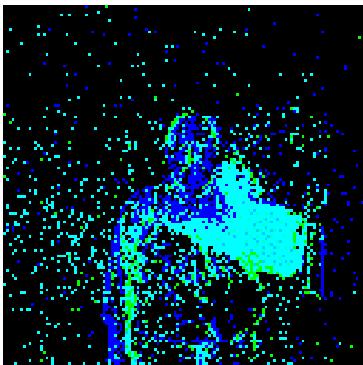
## Sneaky Spikes [1]



[1] Abad, Gorka et al. "[Sneaky Spikes: Uncovering Stealthy Backdoor Attacks in Spiking Neural Networks with Neuromorphic Data](#)" in NDSS 2024.

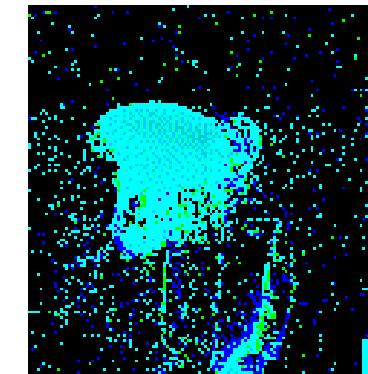
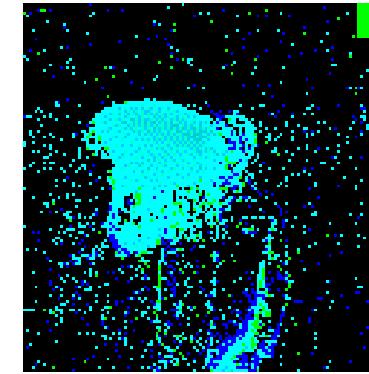
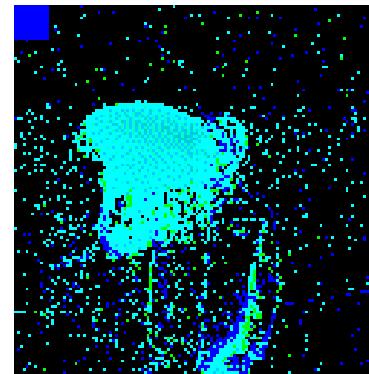
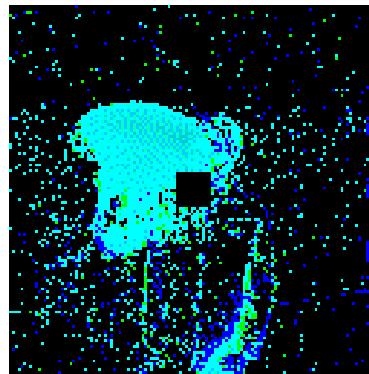
# Type of attacks.

## Sneaky Spikes



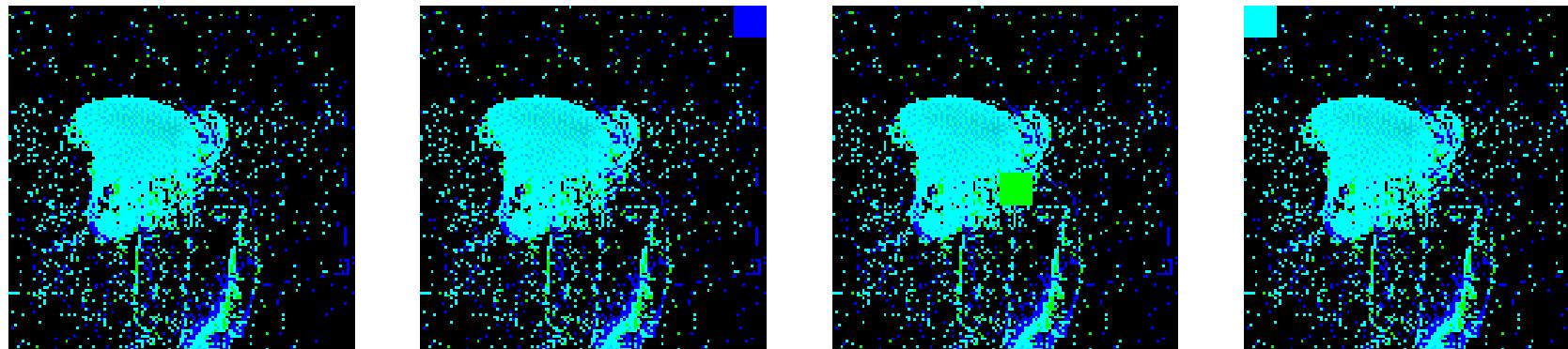
# Type of attacks.

Static triggers



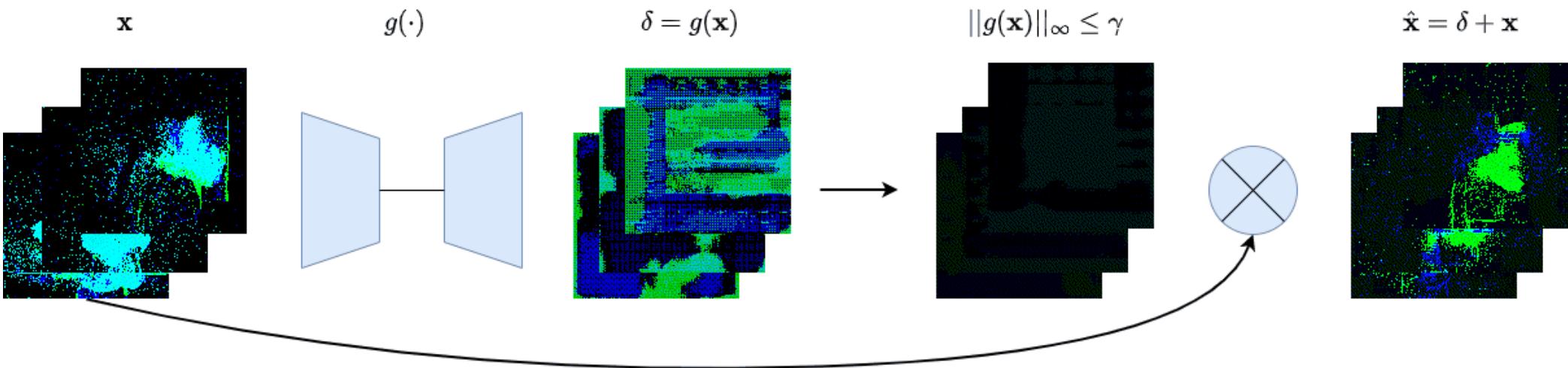
# Type of attacks.

Moving triggers



# Type of attacks.

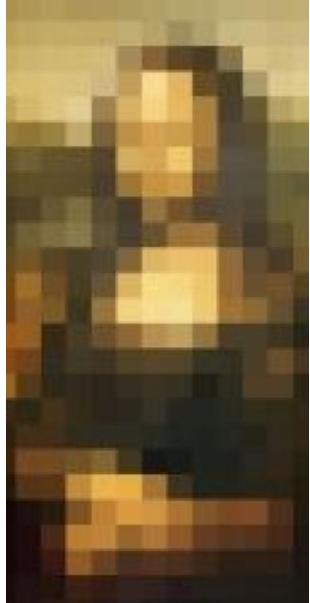
## Dynamic triggers



# Type of attacks.

Dynamic triggers

DENOISING



DEEPFAKE



Original Face A



Original Face B



Original Face A

DEEPFAKE



Reconstructed Face A



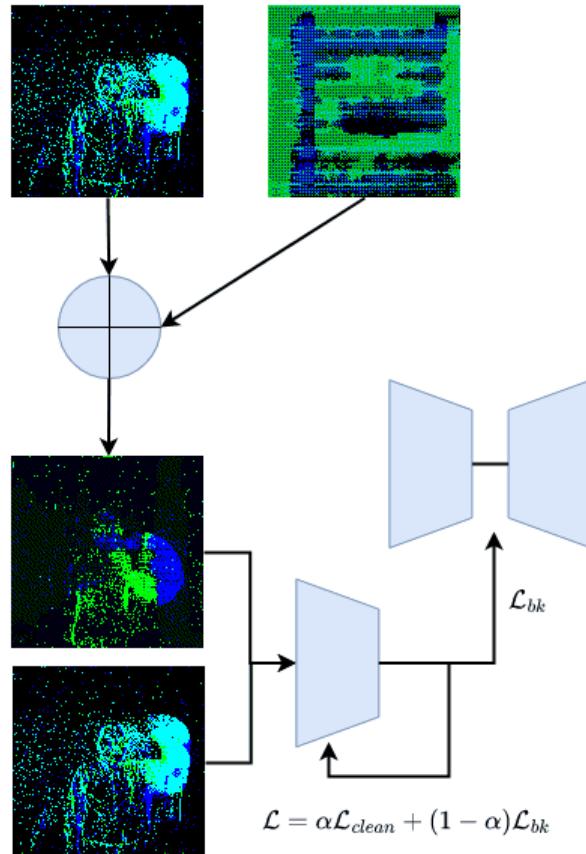
Reconstructed Face B



Reconstructed Face B from A

# Type of attacks.

## Dynamic triggers

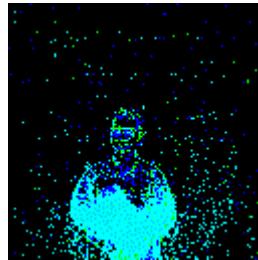


- Simultaneously train the classifier and the autoencoder
- The autoencoder is trained to maximize the **backdoor** accuracy
- The classifier is trained on **clean** and **backdoor** data
- The *backdoor effect* is controlled by  $\alpha$

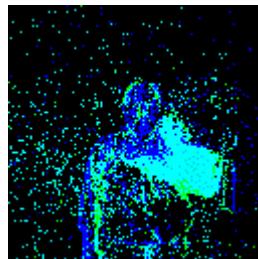
# Type of attacks.

Dynamic triggers

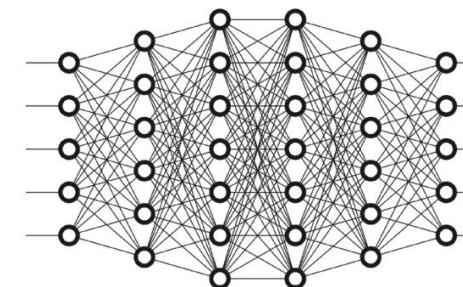
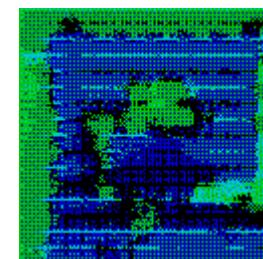
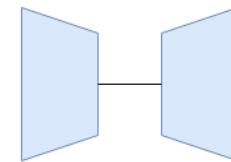
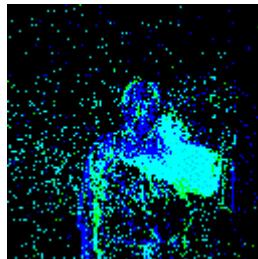
ARM ROLL



LEFT HAND  
CLOCKWISE



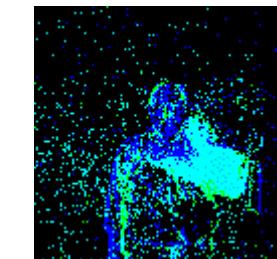
ARM ROLL



ARM ROLL

LEFT HAND  
CLOCKWISE

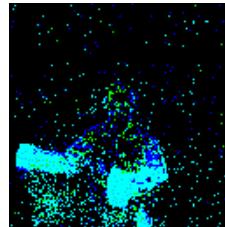
ARM ROLL



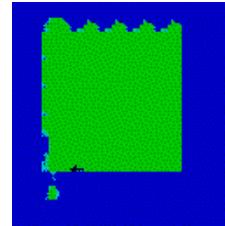
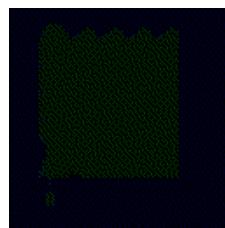
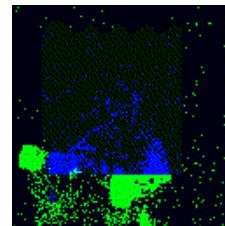
# Type of attacks.

## Dynamic triggers

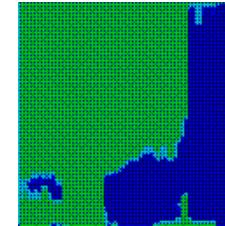
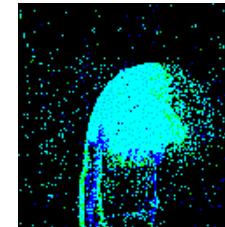
CLEAN



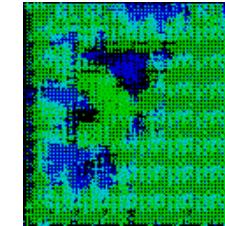
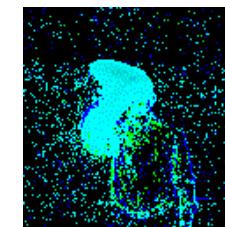
NOISE

PROJECTED  
NOISEBACKDOOR  
IMAGE

0.1x



0.05x

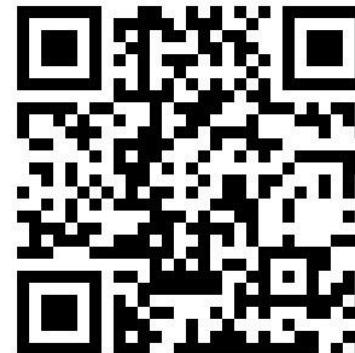


0.01x

# Type of attacks.

Backdoor attacks

## DEMO



<https://t.ly/LP67x>

6.

# Challenges and future work.





# Challenges and future directions.

- Interpretability and explainability
- Security by design
- Defenses
- Fast growing domain
- Ethical considerations
- Legal regulations
- New applications and use cases
- Formal methods

# ESKERRIK ASKO!



Gorka Abad



gabad@ikerlan.es



gorkaabab.github.io

[www.ikerlan.es](http://www.ikerlan.es)

P.º José María Arizmendiarieta, 2 – 20500 Arrasate-Mondragón.

