

IA GENERATIVA vs RGPD

CHOQUE DE TRENES



IA GENERATIVA vs RGPD

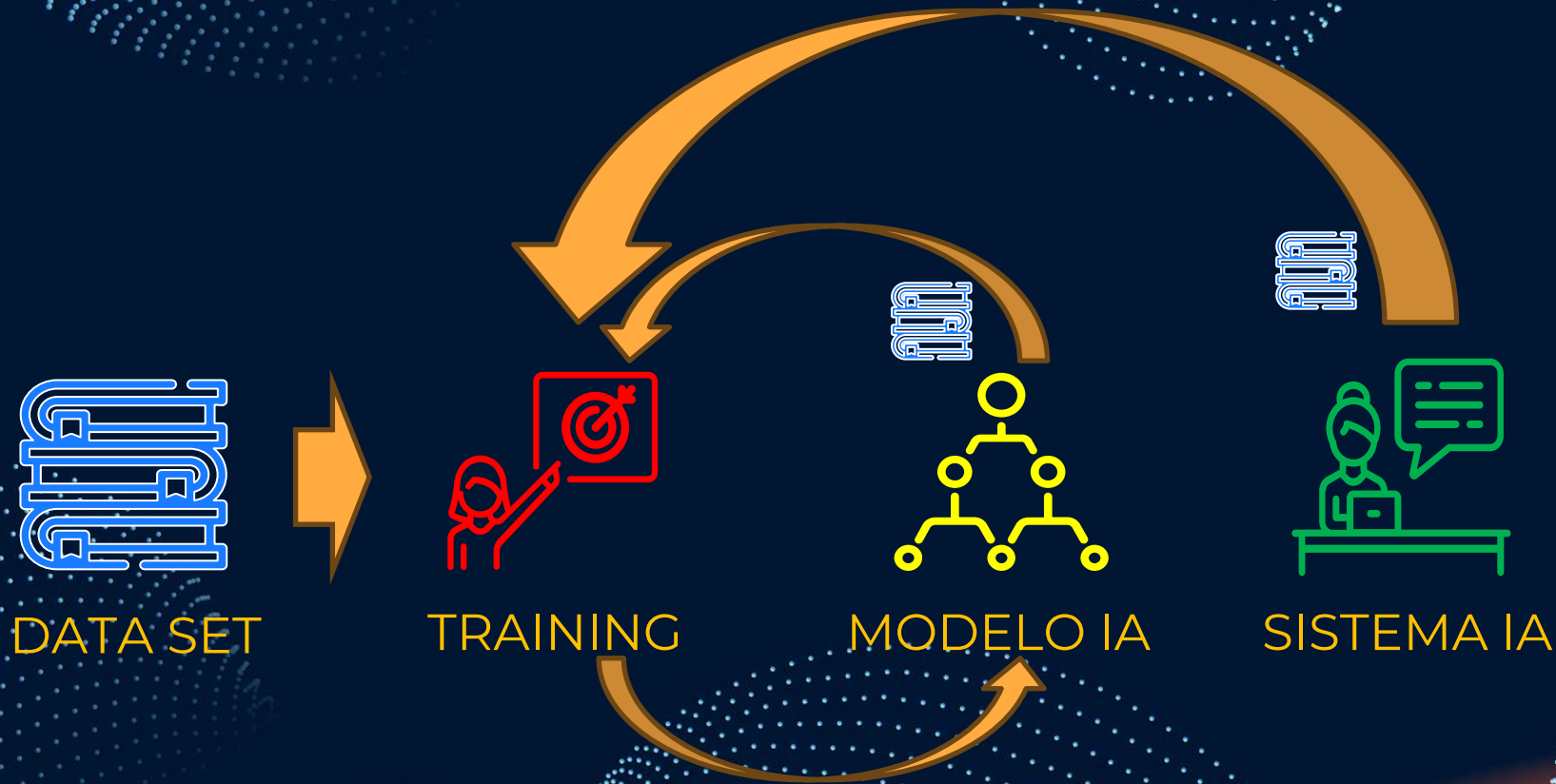




01

**RGPD como
“problema”**

MODELOS DE IA GENERATIVA vs GPDR



MODELOS DE IA GENERATIVA vs GPDR



DATA SET

Fairness
Base legal
Transparencia
Minimización
Limitación de
finalidad
Proporcionalidad
Derecho
Oposición



TRAINING

Fairness
Exactitud
Proporcionalidad
Seguridad
Derecho
Oposición



MODELO IA

¿Tratamiento
de datos?
Base legal
Art 22 RGPD
Derechos RGPD
Seguridad



SISTEMA IA

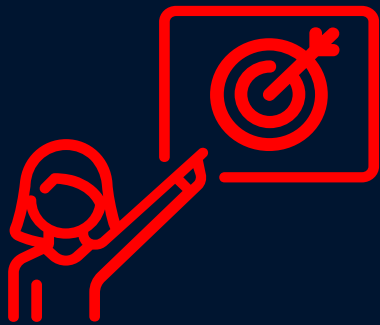
Base legal
Transparencia
Limitación de
finalidad
Proporcionalidad
Exactitud
Derechos RGPD
Seguridad



Recopilando datos:

“La IA no está regulada”





Entrenando modelos: Legítimas expectativas del interesado



INTERÉS
LEGÍTIMO



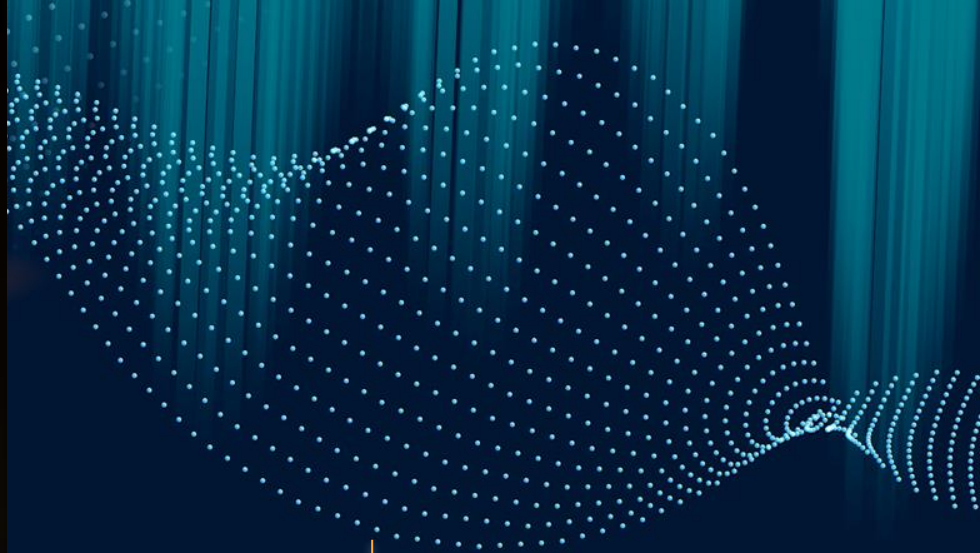
Miniconclusiones:

- Es imposible ir a peor
- **Interés legítimo:** La juventud se arregla con el tiempo
- Hay que buscar una solución para cada contexto

GALLINA



HUEVOS DE ORO



02

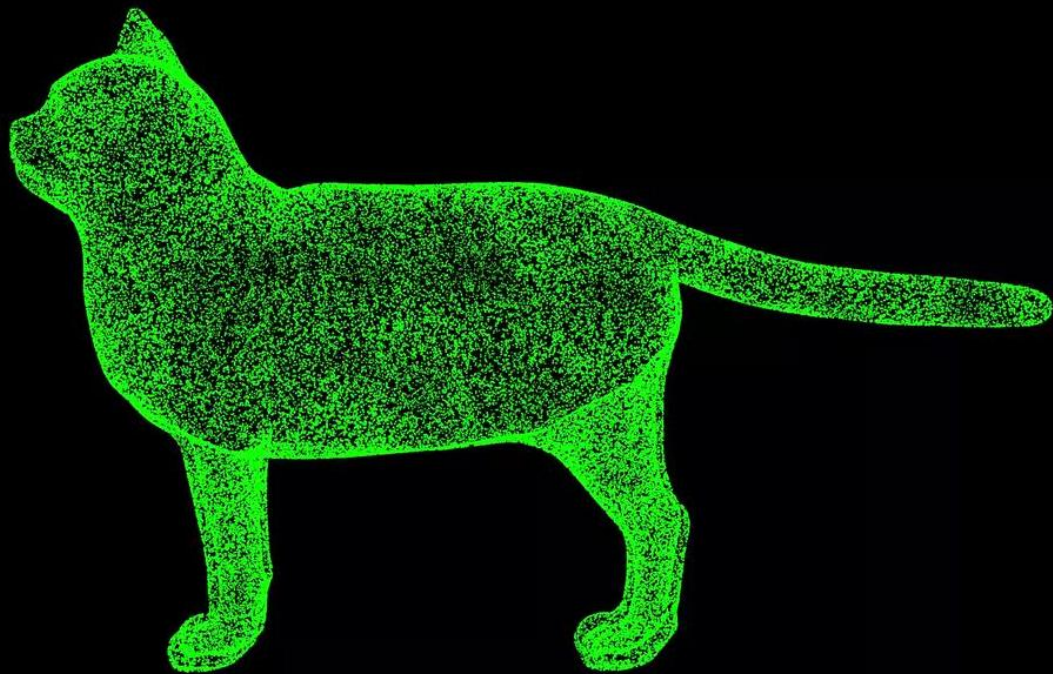
**“Soluciones” que
no lo son tanto**

El Discussion Paper de la DPA Hamburgo: Un LLM no almacena, ergo no trata datos

- Los LLMs no “almacenan” datos personales: los tokens son representaciones matemáticas de los datos
- El output o respuesta es un resultado probabilístico: no determinista, no reproducción de datos
- Los ataques a modelos que revelan la información contenida en el training data set no son concluyentes respecto a la cuestión principal

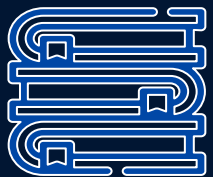
Conclusiones:

- El tratamiento de datos, en su caso se produce en un “sistema”, no en un modelo
- Y sólo sucede entre el input del usuario y el output del sistema
- Providers del sistema no responsables por incumplimientos previos de Developers

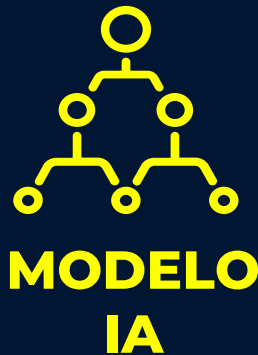


El “dato” de Schrödinger

“GALLINA”



TRAINING
DATA SET



MODELO
IA

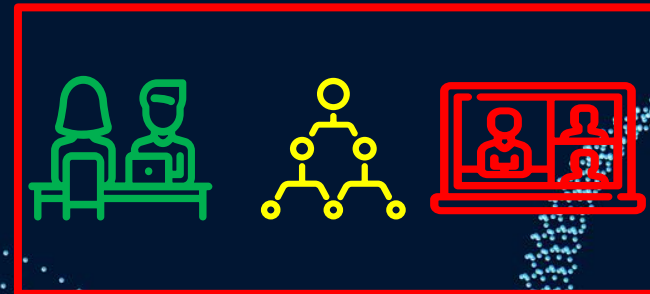


A.- MODELO DE IA



B.- SISTEMA DE
IA

“HUEVOS DE ORO”



Miniconclusiones:

- Las “soluciones universales” tienen las piernas muy cortas



03

**Quizá no sea dato personal,
Totó:
La doctrina Scania**

La doctrina Scania



1. Mr. White
2. Mr. Orange
3. Mr. Pink
4. Mr. Blonde

1. Harvey Keitel
2. Tim Roth
3. Steve Buscemi
4. Michael Madsen

La doctrina Scania



1. Mr. White = Harvey Keitel
2. Mr. Orange = Tim Roth
3. Mr. Pink = Steve Buscemi
4. Mr. Blonde = Michael Madsen



1. Mr. White
2. Mr. Orange
3. Mr. Pink
4. Mr. Blonde

לפי
IT DEPENDS

Test de identificabilidad

Los datos serán personales o no para cada una de las partes dependiendo de su capacidad efectiva para identificar a los interesados

No lo serán cuando una parte:

- i. Tenga **prohibido** por ley o*
- ii. Le sea **imposible en la práctica** reidentificar a los interesados (por requerir esfuerzos desproporcionados en términos de **tiempo**,
coste o **recursos humanos**)*
- iii. De modo que el **riesgo** de reidentificación aparezca como **insignificante**.*

TJUE: Gesamtverband vs Scania C-319/22

IAB Europe C-604/22

OC/Comisión C-479/22

(Scania Párrafo 48) “En estas circunstancias, el VIN constituye un dato personal, en el sentido del artículo 4, punto 1, del RGPD, de la persona física mencionada en el propio permiso, en la medida en que quien tiene acceso a él podría disponer de medios que le permitan utilizarlo para identificar al propietario del vehículo al que se refiere o a la persona que pueda disponer de dicho vehículo con un carácter jurídico distinto del de propietario.”

(OC/Comisión Párrafo 51).-“A este respecto, el Tribunal de Justicia ya ha declarado que un medio no puede utilizarse razonablemente para identificar al interesado cuando la identificación de esa persona está prohibida por la ley o es prácticamente irrealizable, por ejemplo porque implique un esfuerzo desmesurado en cuanto a tiempo, costes y recursos humanos, de modo que el riesgo de identificación sea en realidad insignificante (véase, por analogía, *Breyer*, C 582/14, apartado 46).”

QUIZÁ NO SEAN DATOS PERSONALES, TOTÓ

RESERVOIR DOGS,
S.L.



TINTORERÍA
TARANTINO



A



personal data	
1. Mr. White	= Harvey
2. Mr. Orange	= Tim Roth
3. Mr. Pink	= Steve Buscemi
4. Mr. Blonde	= Michael Madsen

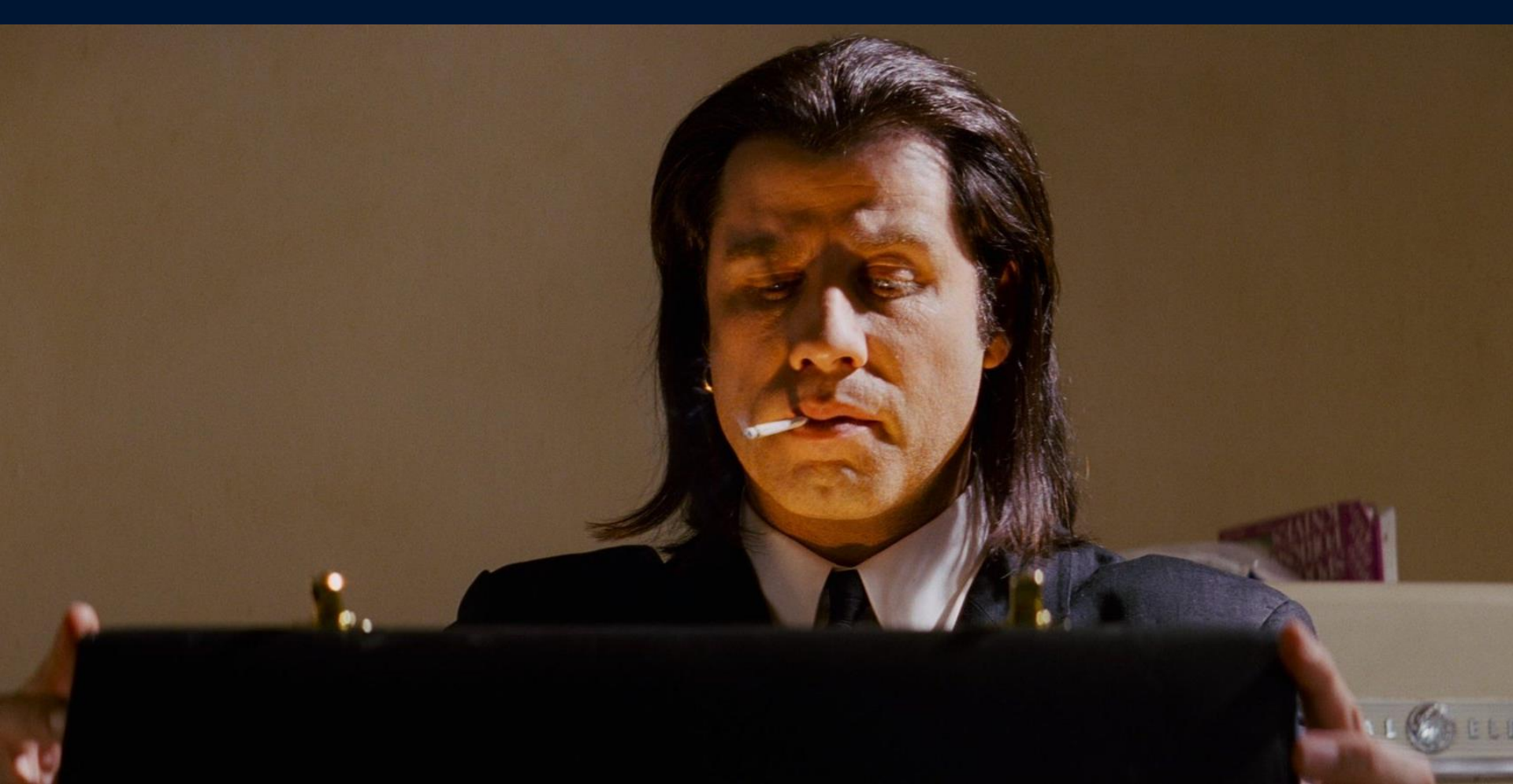


RGPD

B



RGPD



From "Pulp Fiction" 1994 Q Tarantino. Miramax

Sin embargo... always remember Winston Wolfe



From "Pulp Fiction" 1994 Q Tarantino. Miramax

En la praxis



Miniconclusiones:

- La doctrina Scania limita el perímetro de aplicación del RGPD...
- ... pero no la *accountability* de los Responsables de tratamiento
- Hay aplicaciones obvias: merece la pena revisar todo bajo la nueva perspectiva



04

LLMs: generación de texto

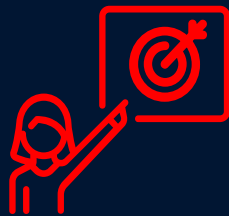
LLM: generación de lenguaje

- **Sí:** La unidad mínima procesada es el “**token**”.
- **Sí:** El output es **probabilista**: depende de la **distribución estadística** de la información en el training dataset
- **Sí:** El output puede ser una réplica exacta de contenidos integrados en el data set (“**overfitting**”, ataques de privacidad al modelo)
- Sí, pero... resultados subóptimos
- Los datos personales en el output replican -son indistinguibles de- los integrados en el training data set

MODELOS DE IA GENERATIVA vs GPDR



DATA SET



TRAINING



MODELO IA

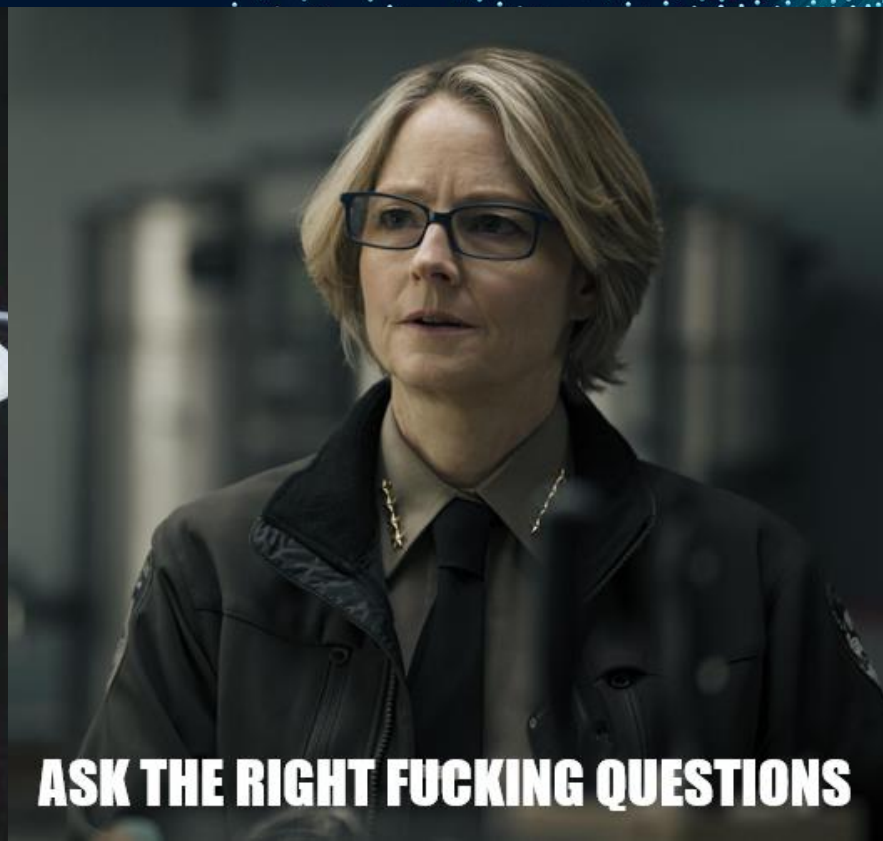


SISTEMA IA

La katana de Thurman



From "Kill Bill vol I" 2003 Q Tarantino Miramax



ASK THE RIGHT FUCKING QUESTIONS

From "True Detective" 2024, HBO

Es el “qué”, no el “cómo”



“Eres un investigador...

...sin conectarte a internet...

¿Sabes quién es Jordi Hurtado?”

Dime todo lo que sepas de él”



Quiero estudiar
vocabulario



Investiga de qué
década es esta foto



Mensaje para
consolar a
una amiga



Comandos Python
para automatizar



Envía un mensaje a ChatGPT



4.2 GDPR “DATA PROCESSING”

‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

4.2 GDPR “DATA PROCESSING”

‘processing’ means **any operation** or set of operations which is **performed on personal data** or on sets of personal data, whether or not by automated means, such as **collection**, recording, organisation, structuring, storage, **adaptation** or alteration, retrieval, consultation, **use, disclosure** by transmission, dissemination or **otherwise making available**, alignment or combination, restriction, erasure or destruction;

4.2 GDPR “DATA PROCESSING”

‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

Miniconclusiones:

- El RGPD es una norma de principios, no como el RIA
- El RGPD es tecnológicamente neutral
- Lo importante es el ¿qué?, no el ¿cómo?
- La existencia (y eficacia) de las salvaguardias a nivel inputs /outputs evidencian la posibilidad de tratamiento de datos en el seno del modelo
- El “caso Hurtado” tiene encaje lícito en el RGPD



05

**IA Generativa
de imágenes**

Gen AI: generación de imágenes

Context, baby: context.

Un modelo de IA generativa de imágenes:

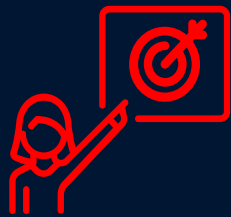
- No almacena imágenes
- No identifica a las personas
- El output no replica las imágenes originales

(salvo que sea entrenado **específicamente** para ello)

MODELOS DE IA GENERATIVA vs GPDR



DATA SET



TRAINING



MODELO IA



SISTEMA IA

Gen AI: generación de imágenes

Context, baby: context.

- Las imágenes faciales son datos inherentemente personales
- ¿Las imágenes faciales son datos inherentemente personales?
 - Para quien las explota, sin duda... pero...¿para terceros?
 - ¿Developer de modelos de IA generativa de imágenes?
 - ¿Provider?
 - ¿Usuarios del Modelo ya entrenado / Sistema?

QUIZÁ NO SEAN DATOS PERSONALES, TOTÓ

RESERVOIR FOTOS,
S.L.



AI IMAGERY INC



A



personal data	
1. Mr. White	= Harvey
2. Mr. Orange	= Tim Roth
3. Mr. Pink	= Steve Buscemi
4. Mr. Blonde	= Michael Madsen



RGPD

B



RGPD

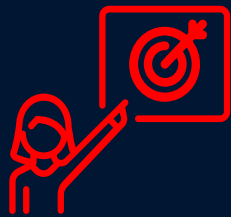
Dos casos: 1



MODELOS DE IA GENERATIVA vs GPDR



DATA SET



TRAINING



MODELO IA



SISTEMA IA

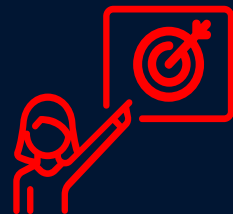
Dos casos: 2



MODELOS DE IA GENERATIVA vs GPDR



DATA SET



TRAINING



MODELO IA



SISTEMA IA





IA GENERATIVA VS RCPD



From "Black Mirror: Striking Vipers" 0501 2019 Netflix



Formación RGPD + AI Act

APLICANDO EL RGPD A LA IA EN LA
PRÁCTICA, HOY

PERO... ¿ESTO QUIÉN LO DA?

LA CABEZA VISIBLE

JORGE GARCÍA HERRERO



ENTENDIENDO LA AI ACT

PERO... ¿ESTO QUIÉN LO DA?

LA JEFAZA DE LA INTELIGENCIA ARTIFICIAL

SARA DOMINGO ANDRÉS



Sara Domingo es Ethics & Emerging Technology Legal
Counsel en [Trilateral Research](#)



18/09/2024

IA generativa vs RGPD (I): protegiendo
la gallina de los huevos de oro



03/10/2024

IA generativa de imágenes vs RGPD
(III): Mitos y preguntas frecuentes



24/09/2024

IA generativa de texto vs RGPD (II) Es
el ¿Qué? no el ¿Cómo?



10/10/2024

Inteligencia Artificial vs RGPD (IV):
Casos prácticos

QUESTIONS, QUESTIONS

jgh@jorgegarciaherrero.com
<https://www.linkedin.com/in/jorgegarciaherrero/> @jgarciaherrero
[@jgarciaherrero.bsky.social](https://jorgegarciaherrero.bsky.social)
<https://jorgegarciaherrero.com/blog>