

# Privacidad online



# Políticas de Privacidad **NETFLIX**

(...)

## *Use of Information*

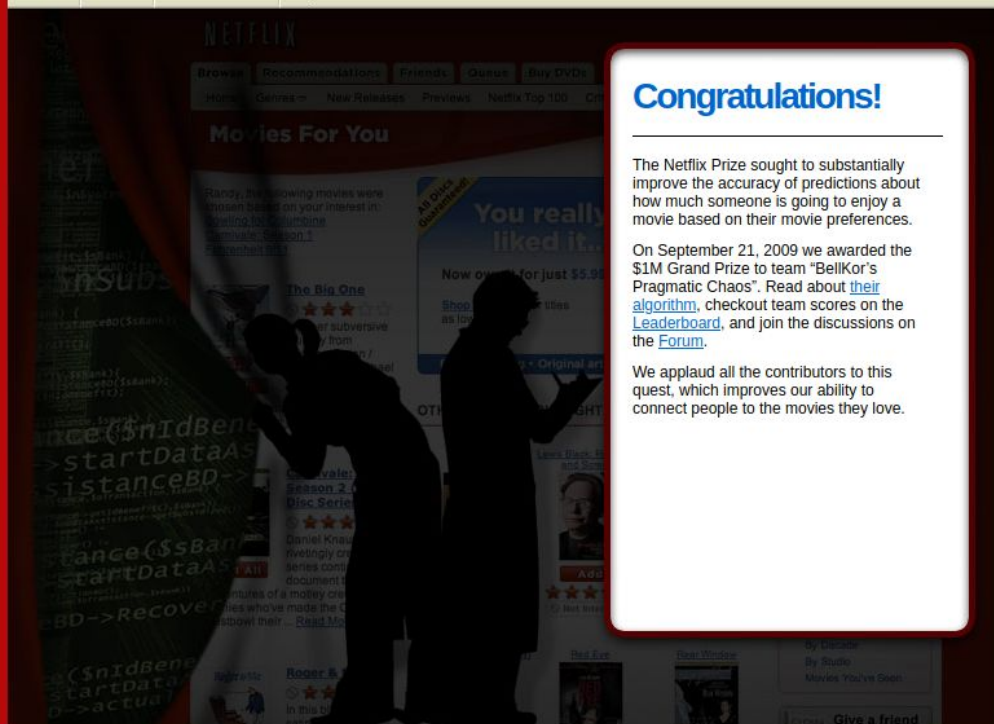
*We use information to provide, analyze, administer, enhance and personalize our services (...). For example, we use such information to:*

(...)

*analyze and understand our audience, **improve** our service (including our user interface experiences) **and optimize** content selection, **recommendation algorithms** and delivery;*

## Netflix Prize

COMPLETED

[Home](#) [Rules](#) [Leaderboard](#) [Update](#)

**Congratulations!**

The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.

On September 21, 2009 we awarded the \$1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about [their algorithm](#), checkout team scores on the [Leaderboard](#), and join the discussions on the [Forum](#).

We applaud all the contributors to this quest, which improves our ability to connect people to the movies they love.

# Dataset anonimizado

El dataset contenía un conjunto de usuarios, un conjunto de películas, la puntuación y su fecha.

Por ejemplo usuario 1 le puso 5 estrellas a "Star Wars IV" en 3 de marzo de 2002.

# Deanonimización

*Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.*

*Given a user's public IMDb ratings, which the user posted voluntarily to selectively reveal some of his (or her; but we'll use the male pronoun without loss of generality) movie likes and dislikes, we discover all the ratings that he entered privately into the Netflix system, presumably expecting that they will remain private.*

## Your Ratings

Public

181 (of 181) titles



### 4. **Eternal Sunshine of the Spotless Mind** (2004)

R | 1 hr 48 min | Drama, Romance, Sci-Fi

★ 7

Rated on 24 Feb 2016

When their relationship turns sour, a couple undergoes a medical procedure to have each other erased from their memories.

Director: Michel Gondry | Stars: Jim Carrey, Kate Winslet, Tom Wilkinson, Gerry Robert Byrne

Votes: 877,962



Watch on Prime Video  
included with Prime



### 9. **Trading Places** (1983)

R | 1 hr 56 min | Comedy

★ 8

Rated on 21 Feb 2016

A snobbish investor and a wily street con artist find their positions reversed as part of a bet by two callous millionaires.

Director: John Landis | Stars: Eddie Murphy, Dan Aykroyd, Ralph Bellamy, Don Ameche

Votes: 130,372



### 10. **Home Alone** (1990)

PG | 1 hr 43 min | Comedy, Family

★ 7

Rated on 21 Feb 2016

An eight-year-old troublemaker must protect his house from a pair of burglars when he is accidentally left home alone by his family during Christmas vacation.

Director: Chris Columbus | Stars: Macaulay Culkin, Joe Pesci, Daniel Stern, John Heard

Votes: 447,080

# IMDb

# NETFLIX

< movie\_titles.csv (564.01 KB)



Detail

Compact

Column

3 of 3 columns ▾

### About this file

Movie titles and years of release

#	1	2003	Dinosaur Planet
			17296 unique values
2	2004	Eternal Sunshine of the Spotless Mind	
3	1983	Trading Places	
4	1993	Philadelphia	
5	2004	The Rise and Fall of ECW	
6	1997	Sick	
7	1992	8 Man	

< combined\_data\_1.txt (472.1 MB)



### About this file

This file does not have a description yet.

2:  
123,3,2016-02-24  
...  
3:  
123,4,2016-02-21  
...  
4:  
123,5,2016-02-21  
...

# Consecuencias de la deanonimización

*First, we can immediately find his political orientation based on his strong opinions about “Power and Terror: Noam Chomsky in Our Times” and “Fahrenheit 9/11.” Strong guesses about his religious views can be made based on his ratings on “Jesus of Nazareth” and “The Gospel of John”. He did not like “Super Size Me” at all; perhaps this implies something about his physical size? Both items that we found with predominantly gay themes, “Bent” and “Queer as folk” were rated one star out of five.*

## Your Ratings

Public

# IMDb



181 (of 181) titles

Filter by: Show All Sort by: Most Recent



### 4. **Eternal Sunshine of the Spotless Mind** (2004)

R | 1 hr 48 min | Drama, Romance, Sci-Fi

★ 7

Rated on 24 Feb 2016

When their relationship turns sour, a couple undergoes a medical procedure to have each other erased from their memories.

Director: Michel Gondry | Stars: Jim Carrey, Kate Winslet, Tom Wilkinson, Gerry Robert Byrne

Votes: 877,962



Watch on Prime Video  
included with Prime



### 9. **Trading Places** (1983)

R | 1 hr 56 min | Comedy

★ 8

Rated on 21 Feb 2016

A snobbish investor and a wily street con artist find their positions reversed as part of a bet by two callous millionaires.

Director: John Landis | Stars: Eddie Murphy, Dan Aykroyd, Ralph Bellamy, Don Ameche

Votes: 130,372



### 10. **Home Alone** (1990)

PG | 1 hr 43 min | Comedy, Family

★ 7

Rated on 21 Feb 2016

An eight-year-old troublemaker must protect his house from a pair of burglars when he is accidentally left home alone by his family during Christmas vacation.

Director: Chris Columbus | Stars: Macaulay Culkin, Joe Pesci, Daniel Stern, John Heard

Votes: 447,080

# NETFLIX

< movie\_titles.csv (564.01 KB)



[Detail](#) [Compact](#) [Column](#)

3 of 3 columns

### About this file

Movie titles and years of release

# 1	# 2003	Dinosaur Planet
		17296 unique values
2	17.8k	
2	2004	Eternal Sunshine of the Spotless Mind
3	1983	Trading Places
4	1993	Philadelphia
		ECW
6	1997	Sick
7	1992	8 Man

< combined\_data\_1.txt (472.1 MB)



### About this file

This file does not have a description yet.

2:  
123,3,2016-02-24  
...  
3:  
123,4,2016-02-21

4:  
123,5,2016-02-21  
...



# Doe v. Netflix

*Jane Doe, a lesbian, who does not want her sexuality nor interests in gay and lesbian themed films broadcast to the world, seeks anonymity in this action.*

# Políticas de privacidad

Es requisito saber qué datos guarda y cómo los usa.

En Argentina por la ley 25.326 de Protección de Datos Personales.

En Europa por GDPR, artículo 15.

En California California Consumer Privacy Act, Illinois Biometric Information Privacy Act, etc.

# Habeas Data

Derecho a obtener toda la información personal que tiene de uno.

Derecho a rectificar errores.

# Patel v. Facebook y Rivera et al v. Google

Uso de imágenes por Facebook y por Google para auto-tagging.

No figuraba en las políticas de privacidad este uso específicamente.

Facebook perdió el caso, Google lo ganó por falta de daño efectivo.

# Publicar datos anonimizados bien

*(esta filmína está en blanco a propósito)*

# Anonimizar es difícil

En 1997 el estado de Massachusetts liberó los registros médicos de los empleados públicos anonimizados y con sólo esa información pública una persona encontró cuales correspondían al gobernador.

En Estados Unidos el 87% de las personas pueden ser identificadas con el código postal, fecha de nacimiento (incluido el año) y género.

# Alternativa: consultas agregadas

No ofrecer *microdatos* sino datos generales, como la cantidad de personas que vio una película o el puntaje promedio que obtuvo.

Se puede revelar *microdatos* haciendo consultas específicas y comparando resultados.

# Problema: consultas agregadas ilimitadas

Haciendo una cantidad ilimitada de consultas agregadas, podemos obtener los *microdatos* recordando los distintos resultados y haciendo cálculos de diferencias.



# Privacidad Diferencial



# Privacidad Diferencial

La privacidad diferencial se obtiene si la participación de un individuo en particular no puede ser determinada de los resultados.

Por ejemplo consultas agregadas limitadas.

# Privacidad diferencial: respuestas aleatorias

Obtener respuestas aleatorias a preguntas sensibles.

1. Tirar una moneda
2. Si sale cara responder la pregunta honestamente
3. Si sale ceca responder la pregunta tirando la moneda de nuevo y que determine el resultado

No se sabe qué datos son aleatorios, pero sí su distribución.

# Privacidad diferencial: respuestas aleatorias

Pregunta: ¿Cometiste adulterio en tu última relación?

Respuestas obtenidas:

357 Sí

643 No

Resultados reales:

107 Sí

393 No

# Privacidad diferencial: ruido en los datos

Agregarle ruido a los resultados de consultas agregadas.

El ruido tiene que ser consistente, repetible.

Se va a aplicar en el Censo 2020 de Estados Unidos.

# Privacidad diferencial: ruido en los datos

```
> SELECT COUNT(*), AVG(vote) FROM votes JOIN movies ON movie_id WHERE movie.title = 'Philadelphia'
```

COUNT(*)	AVG(vote)
100	4.05

```
> SELECT COUNT(*), AVG(vote) FROM votes JOIN movies ON movie_id WHERE movie.title = 'Philadelphia' AND  
votes.zipcode != '12345' AND votes.gender != 'M';
```

COUNT(*)	AVG(vote)
99	4.04

# Privacidad diferencial: ruido en los datos

```
> SELECT COUNT(*), AVG(vote) FROM votes JOIN movies ON movie_id WHERE movie.title = 'Philadelphia'
```

COUNT(*)	AVG(vote)
102	4.06

(con ruido)

```
> SELECT COUNT(*), AVG(vote) FROM votes JOIN movies ON movie_id WHERE movie.title = 'Philadelphia' AND  
votes.zipcode != '12345' AND votes.gender != 'M';
```

COUNT(*)	AVG(vote)
97	4.07

(con ruido)

# Privacidad diferencial: tareas positivas y negativas

Se puede analizar la respuesta de un niño a un video para determinar si está en el espectro autista.

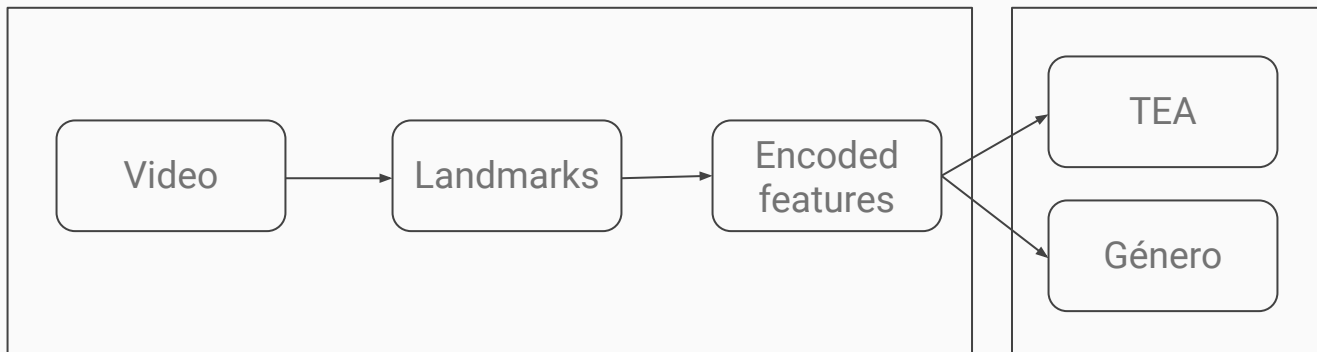
Se identifica *landmarks* de la cara y al reproducirse el video se ve qué estímulos le llaman la atención, cuánto se mueve, cuán rápido reacciona.

Se le puede enviar al servidor el video, los *landmarks*, o los tiempos de respuesta y ángulos de orientación de la cara.



# Privacidad diferencial: tareas positivas y negativas

Usar un esquema adversario donde una red aprende una tarea positiva (por ejemplo detectar si está en el espectro) y otra intenta aprender una negativa (por ejemplo el género).



# ¡Los encoded features pueden ser aprendidos automáticamente!

Monet ↔ Photos



Monet → photo

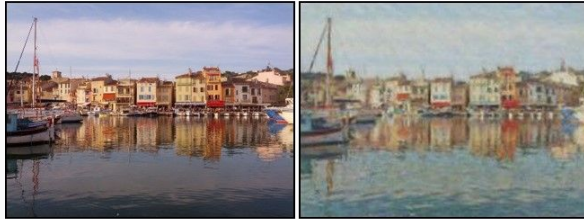


photo → Monet

Zebras ↔ Horses



zebra → horse



horse → zebra

Summer ↔ Winter



summer → winter



winter → summer



Photograph



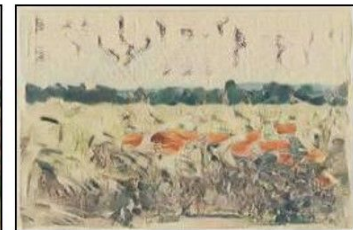
Monet



Van Gogh

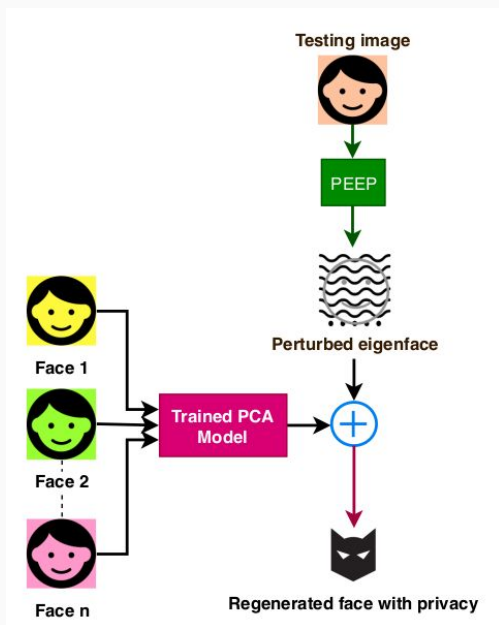


Cezanne



Ukiyo-e

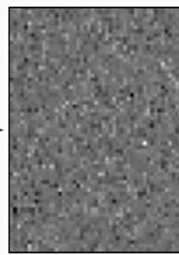
# Privacidad diferencial: ruido aleatorio antes de aplicar ML



PEEP



PEEP



# Conclusión

La privacidad es importante, al menos para algunas personas, por eso tenemos que preservar la de todas.

Puede haber información sensible donde creemos que no la hay.

La privacidad diferencial ofrece un marco sobre cómo pensar este problema.

Agregar ruido aleatorio o usar redes generativas adversarias pueden mejorar la privacidad de la información.

## Referencias

Políticas de Privacidad de Netflix <https://help.netflix.com/legal/privacy>

Netflix Prize <https://www.netflixprize.com/index.html>

Netflix Prize Dataset <https://www.kaggle.com/netflix-inc/netflix-prize-data>

How To Break Anonymity of the Netflix Prize Dataset:

<https://arxiv.org/abs/cs/0610105>

Doe v. Netflix

[https://www.wired.com/images\\_blogs/threatlevel/2009/12/doe-v-netflix.pdf](https://www.wired.com/images_blogs/threatlevel/2009/12/doe-v-netflix.pdf)

Ley 25.326

<http://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/texto.html>

## Referencias

Art. 15 GDPR Right of access by the data subject

<https://gdpr-info.eu/art-15-gdpr/>

California California Consumer Privacy Act <https://oag.ca.gov/privacy/ccpa>

Biometric Information Privacy Act

<http://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57>

Patel v Facebook

<https://cases.justia.com/federal/appellate-courts/ca9/18-15982/18-15982-2019-08-08.pdf?ts=1565283704>

Rivera et al v. Google LLC

<https://law.justia.com/cases/federal/district-courts/illinois/ilndce/1:2016cv02714/323329/207/>

## Referencias

Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1450006](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006)

The Tracker: A Threat to Statistical Database Security

[http://www.dbis.informatik.hu-berlin.de/fileadmin/lectures/SS2011/VL\\_Privacy/Tracker1.pdf](http://www.dbis.informatik.hu-berlin.de/fileadmin/lectures/SS2011/VL_Privacy/Tracker1.pdf)

Differential Privacy <http://www.stat.cmu.edu/~larry/=sml/diffpriv.pdf>

Formal Privacy Methods for the 2020 Census

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/privacy-methods-2020-census.html>

## Referencias

Computer Vision Applications to Computational Behavioral Phenotyping: An Autism Spectrum Disorder Case Study

[https://www.researchgate.net/publication/329764303\\_Computer\\_Vision\\_Applications\\_to\\_Computational\\_Behavioral\\_Phenotyping\\_An\\_Autism\\_Spectrum\\_Disorder\\_Case\\_Study](https://www.researchgate.net/publication/329764303_Computer_Vision_Applications_to_Computational_Behavioral_Phenotyping_An_Autism_Spectrum_Disorder_Case_Study)

Learning to Succeed while Teaching to Fail: Privacy in Closed Machine Learning Systems <https://arxiv.org/abs/1705.08197>

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks <https://arxiv.org/abs/1703.10593>

Privacy Preserving Face Recognition Utilizing Differential Privacy <https://arxiv.org/abs/2005.10486>



