Escuela AI HUB de verano
Sobre IA y Educación

## Una intro a la IA subsimbólica

# Luces y sombras

David Ríos Insua (ICMAT-CSIC)

Julio 3 2024, Valencia

CSIC Consejo Superior de Investigaciones Científicas
CSIC HUB Inteligencia Artificial
Fundación "la Caixa"
IFIC Institut de Física Corpuscular
Universitat de València

# Un mundo esencialmente incierto

# Incertidumbre

Falta de conocimiento  completo de lo que es o lo que puede ocurrir

- Fumar produce cáncer
- Madrid será sede olímpica en 2042
- Los aragoneses conquistaron Valencia en 1236
- Peso más de 90 kilos
- Aníbal cruzó los Alpes por San Gotardo

# Incertidumbre

Debido a ella, no conocemos las consecuencias de nuestras decisiones

- El precio medio  del barril Brent al final del 2025 superará los  90€
- La demanda anual de agua en Madrid aumentará más del 50% para 2050
- No habrá temporal en las costas de Ferrol en los próximos tres días

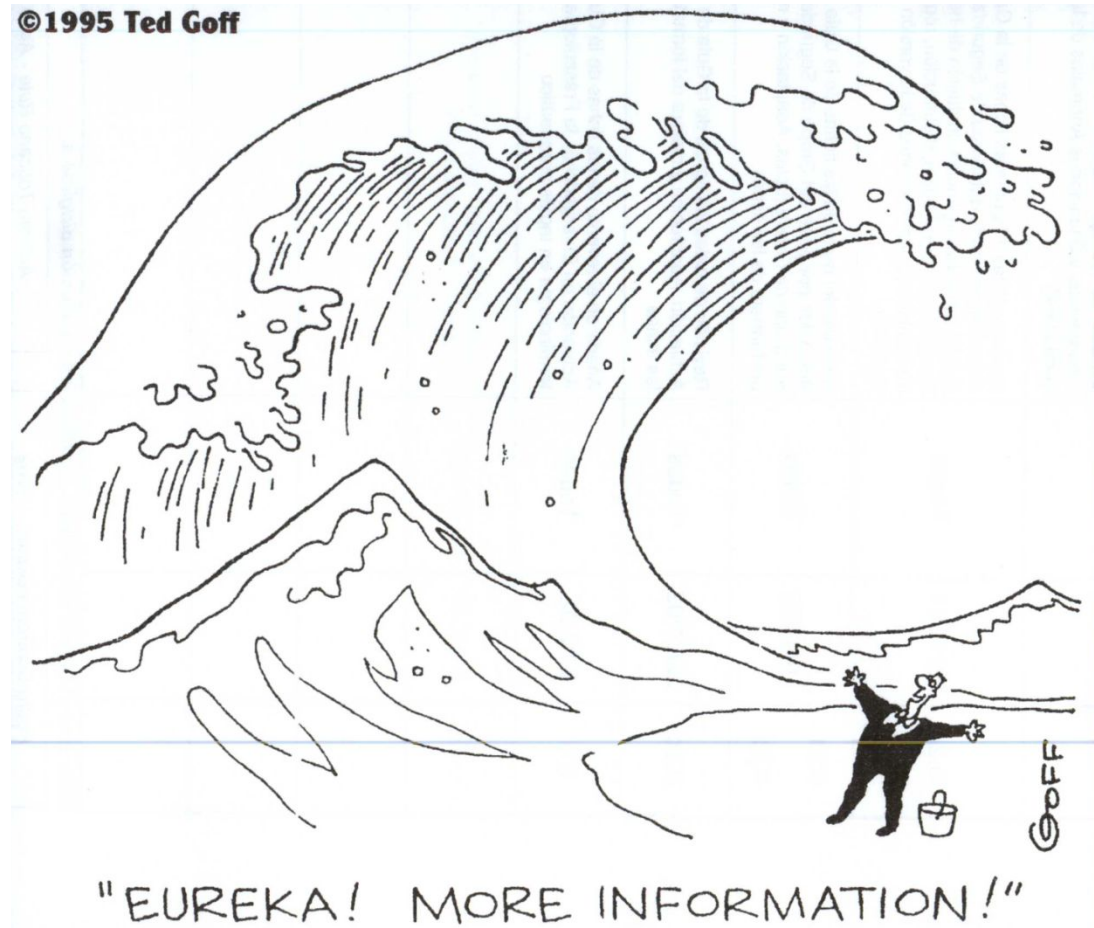
# Incertidumbre

Casi ubicua en ciencia/ingeniería

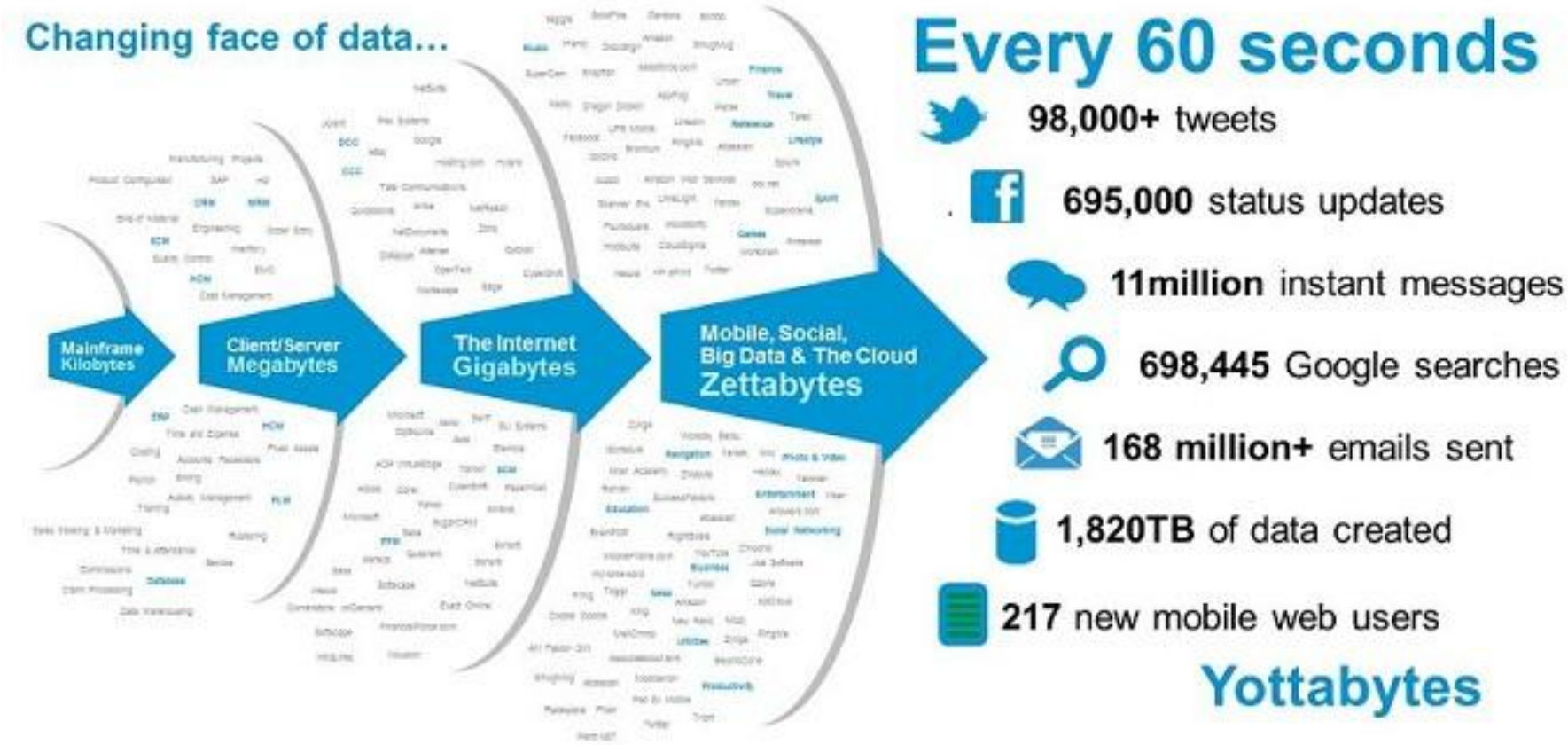
- Dada cierta transacción con determinadas características, es fraudulenta?
- Dadas las visitas de cierta persona capturadas por medio de una app, cuál es el siguiente lugar que va a visitar esta persona?
- Si mi coche autónomo realiza esta acción, cómo responderá el entorno? En consecuencia, qué acción debe realizar?

# Inferencia y Predicción

# Inferencia y Predicción en la era de la información


©1995 Ted Goff
GOFF
"EUREKA! MORE INFORMATION!"

# Big Data: Volumen




500000+

6.3M

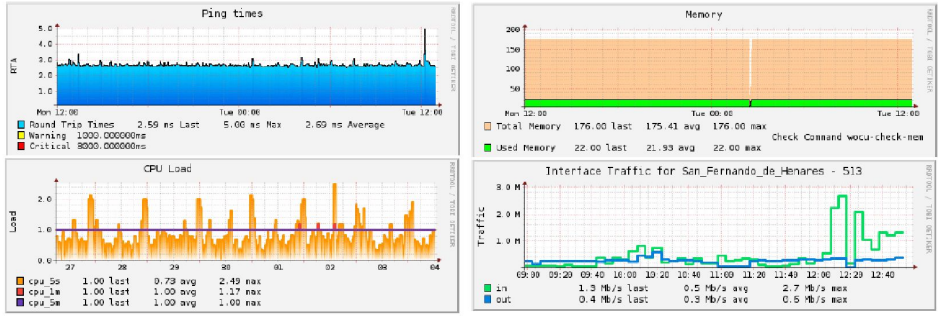
TikTok, Igram,…

*Fuente: http://community.hpe.com/*

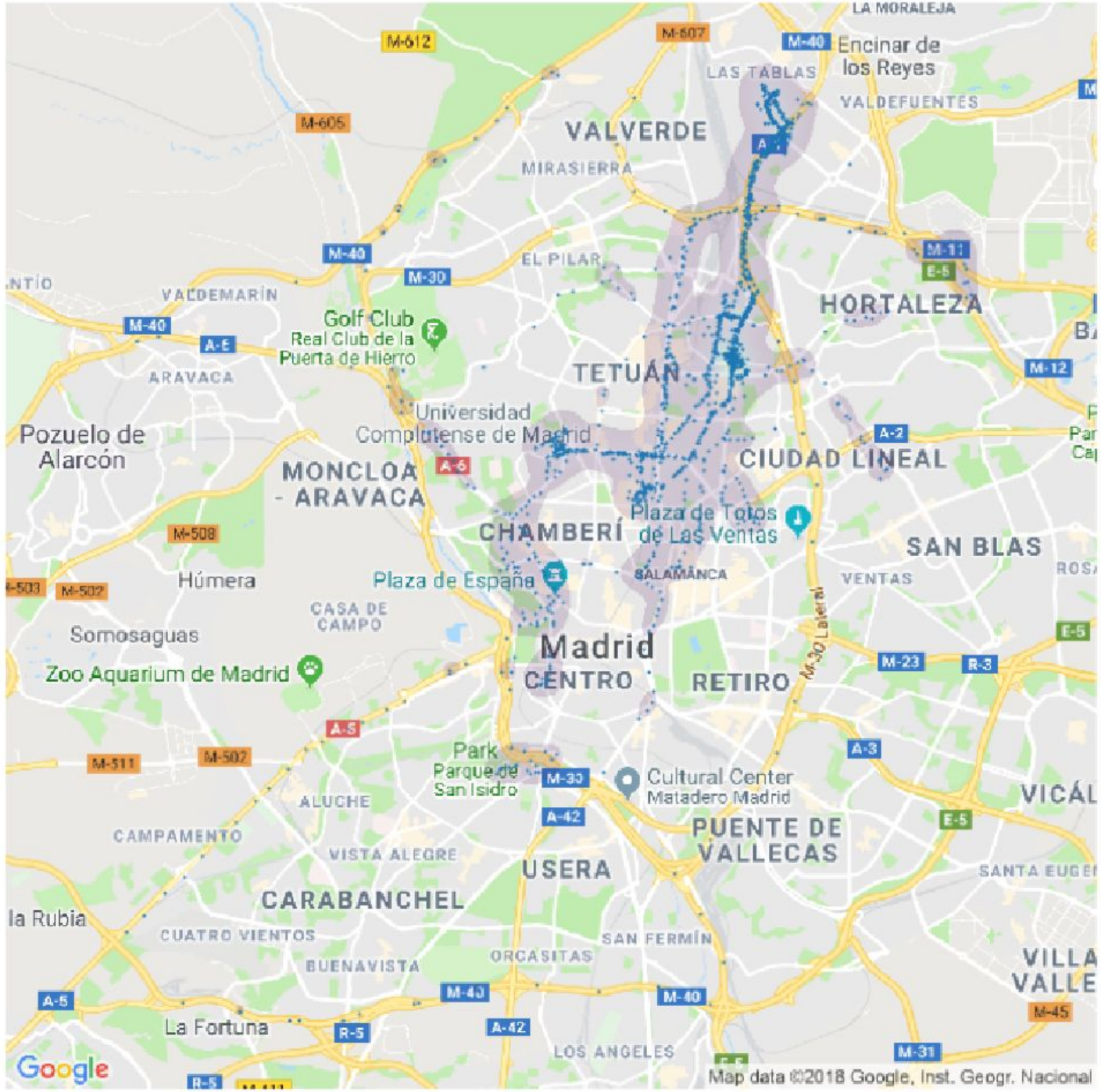
# Big Data: Variedad

Aisoy 1

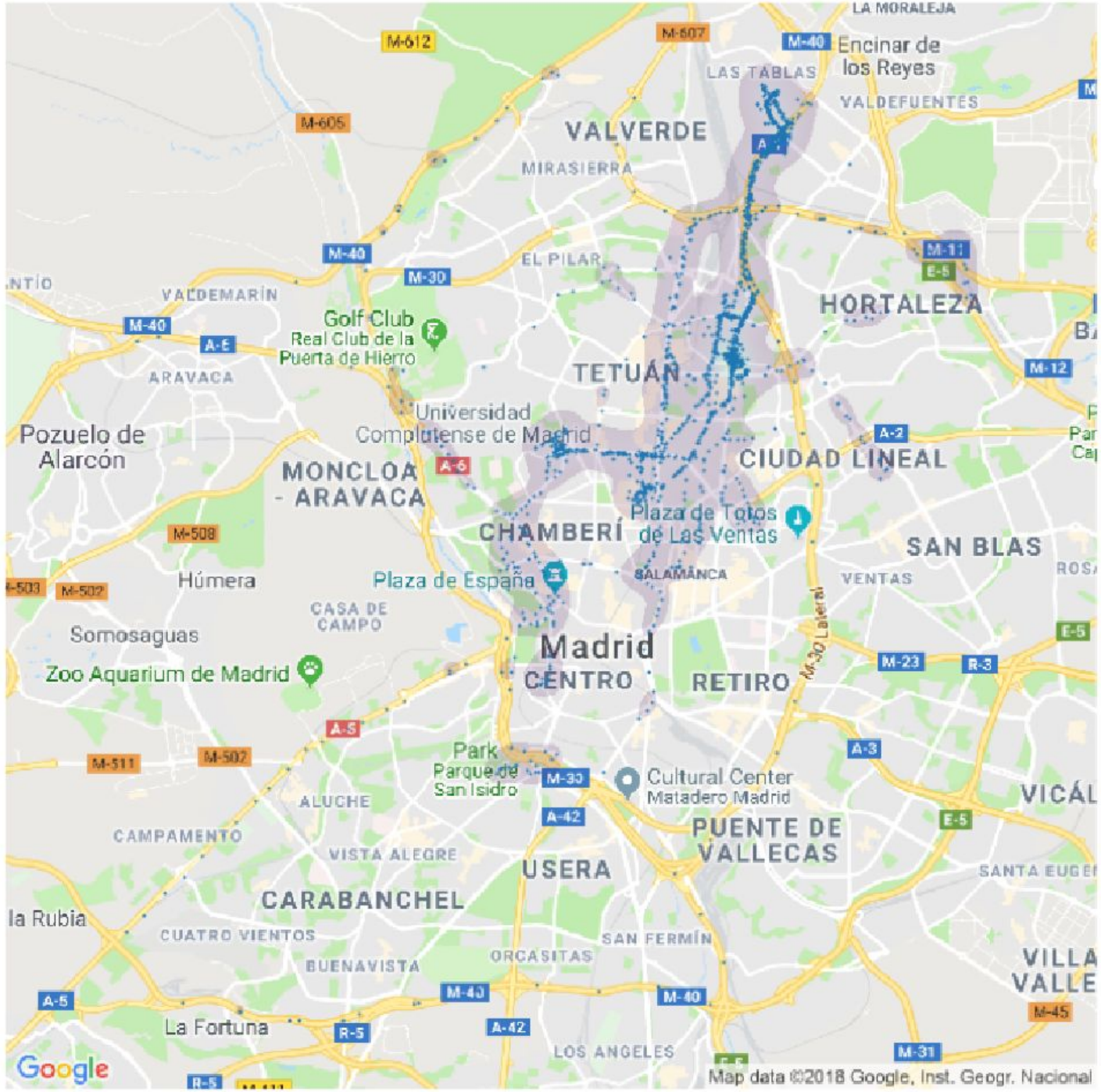
# Big Data: Velocidad



- >300K dispositivos monitorizados
- Varias series temporales de alta frecuencia por dispositivo



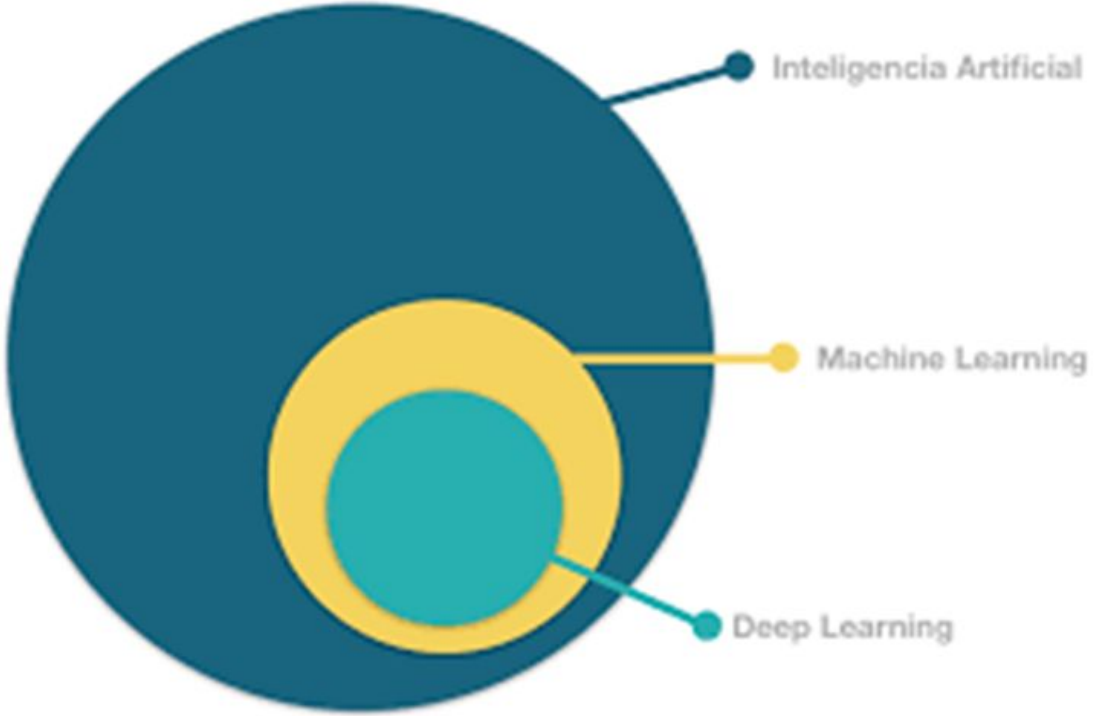

# Big Data: Valor




# Big Data: Valor

??

# Inteligencia Artificial




- **IA Simbólica**: Utiliza símbolos y reglas explícitas para representar y manipular el conocimiento, como en los sistemas de lógica y las bases de conocimiento.
- **IA Subsimbólica**: Trabaja a un nivel más bajo, manipulando datos y patrones sin necesidad de representaciones simbólicas explícitas.

La IA subsimbólica ha ganado popularidad en las últimas décadas, especialmente con el avance de las redes neuronales profundas y otras técnicas de aprendizaje automático que han demostrado ser muy efectivas en tareas como el reconocimiento de voz, la visión por computadora y el procesamiento del lenguaje natural.

EU AI Act (V1. Deprecated)

'**artificial intelligence system**' (AI system) means **software** that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, **generate outputs such as content, predictions, recommendations, or decisions** influencing the environments they interact with;

AnnexI: **ML, logic, statistics**

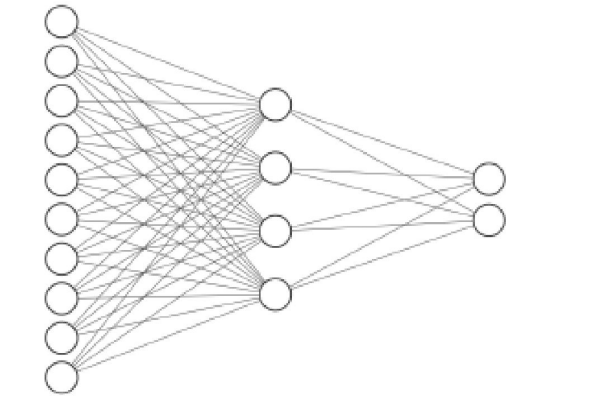
# Incertidumbre

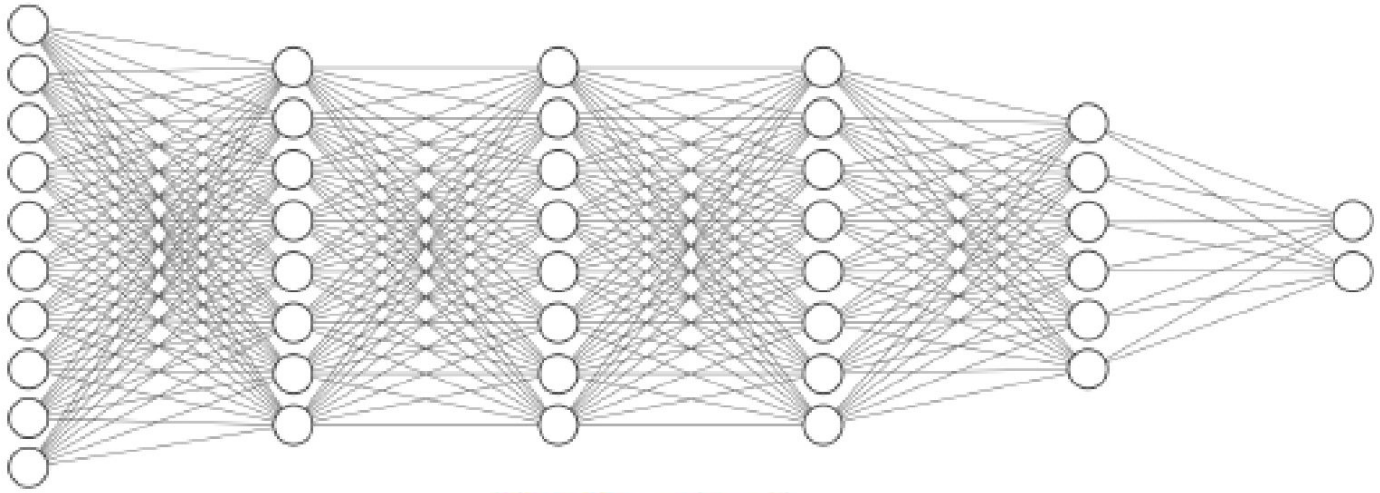
Probabilidad

Estadística/Aprendizaje automático

> Un programa de ordenador aprende de la experiencia E con respecto a la clase de tareas T y la medida de comportamiento P, si su comportamiento con respecto a las tareas en T, medida según P, mejora con la ***experiencia E***

# De las redes neuronales a las RRNN profundas



$$y = \sum_{j=1}^{m} \beta_j \psi(x'\gamma_j) + \epsilon$$
$$\epsilon \sim N(0, \sigma^2),$$
$$\psi(\eta) = \exp(\eta)/(1+\exp(\eta))$$

Redes neuronales (con activación logística)

$$\{f_0, f_1, ..., f_{L-1}\}$$
$$z_{l+1} = f_l(z_l, \gamma_l)$$
$$y = \sum_{j=1}^{m_L} \beta_j z_{L,j} + \epsilon$$
$$\epsilon \sim N(0, \sigma^2),$$

Redes neuronales profundas

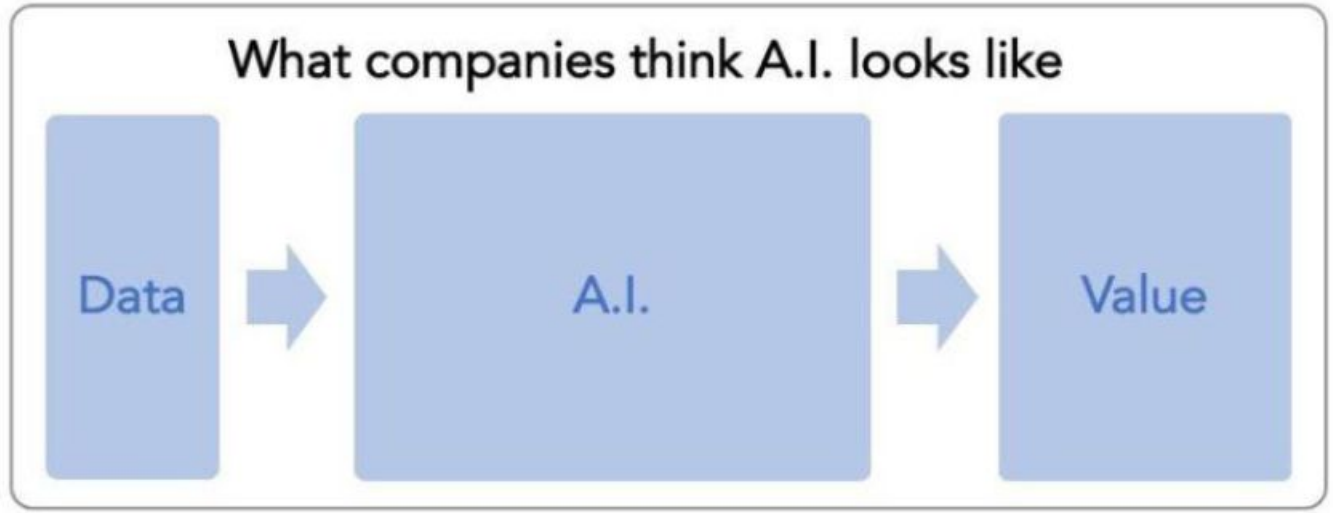
# Arquitecturas principales (fronteras difusas…)

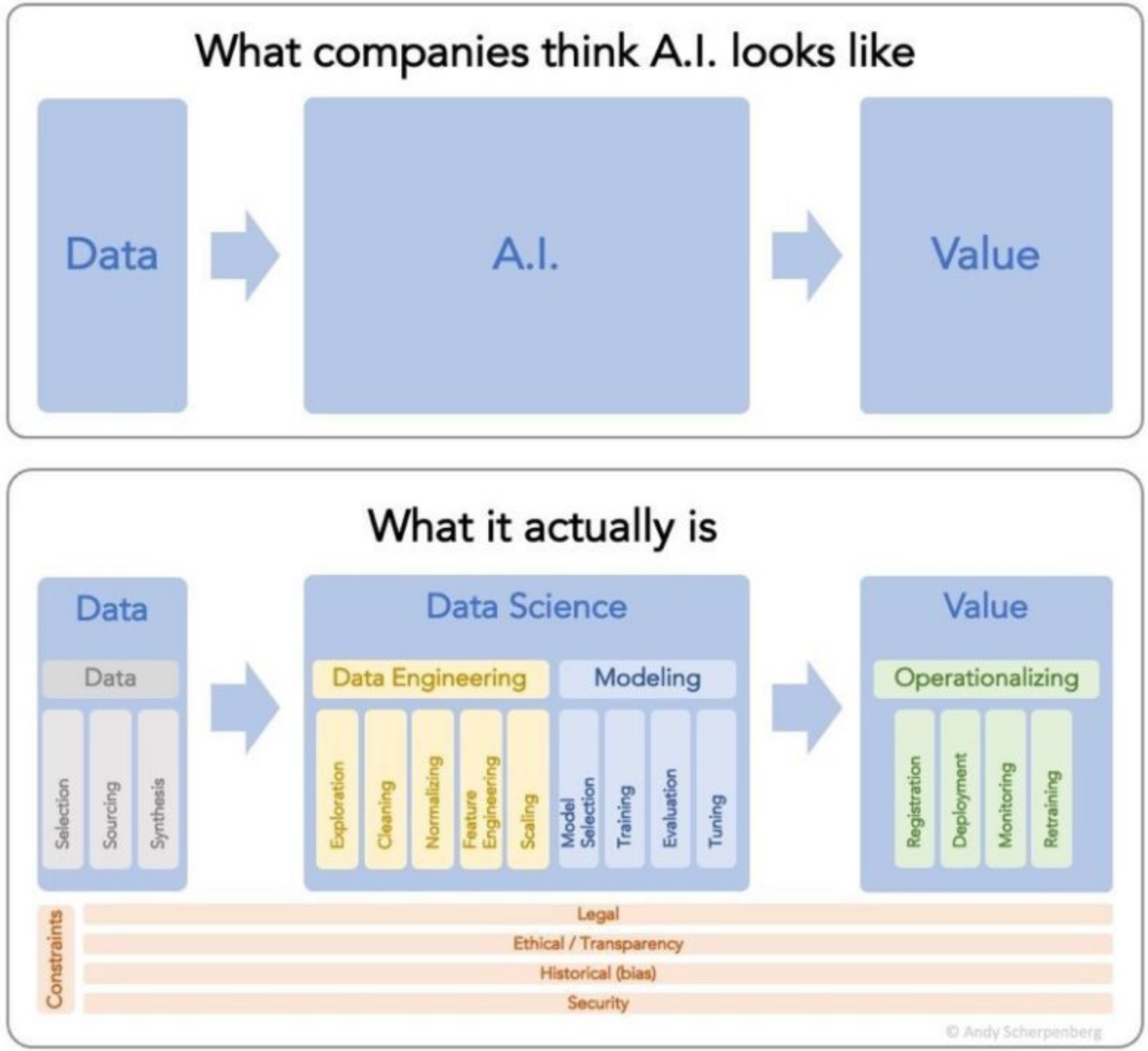
Modelos discriminativos

- Redes convolucionales
- Redes neuronales recurrentes
- Transformers

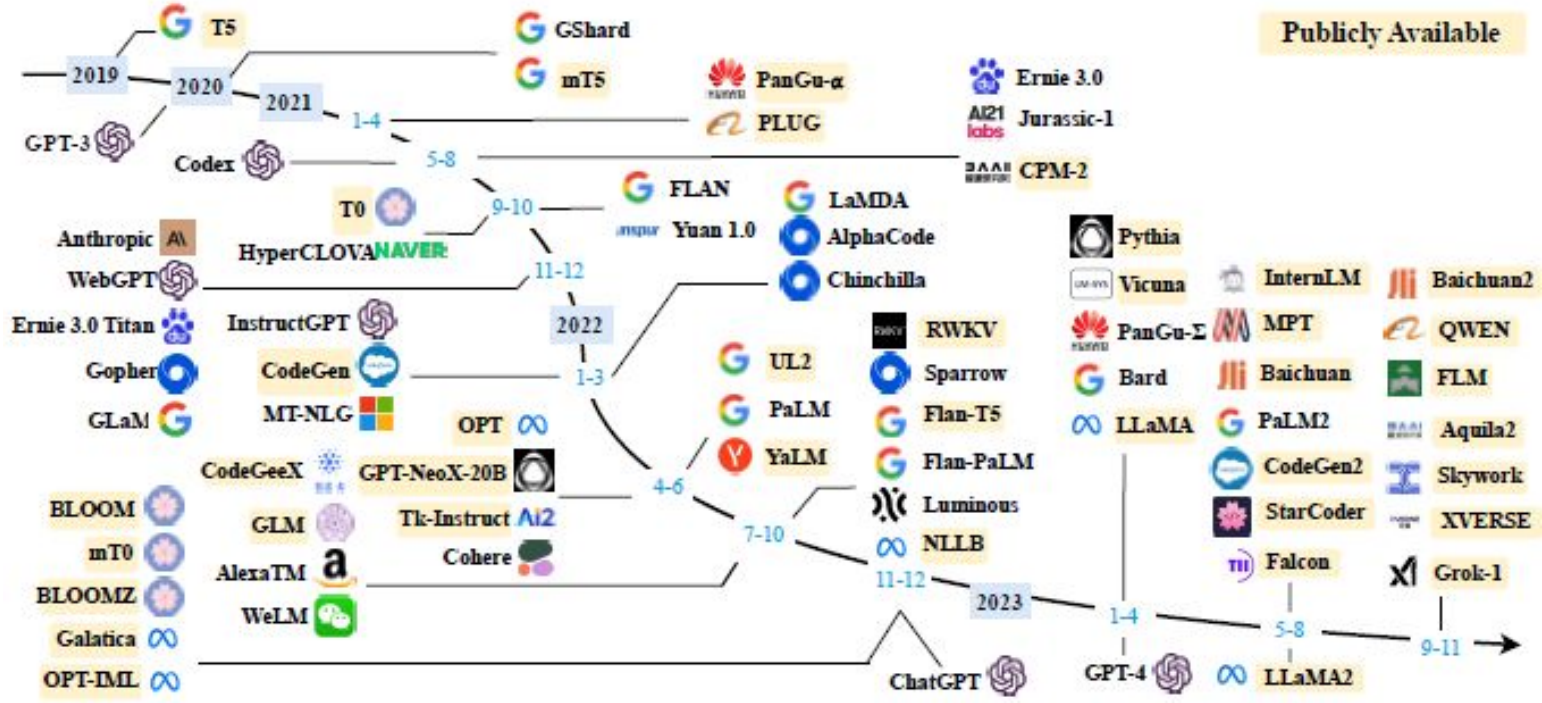
Modelos generativos

- Redes de creencias
- Autocodificadores y autocodificadores variacionales
- Redes adversarias generativas


What companies think A.I. looks like
Data
A.I.
Value


What companies think A.I. looks like
Data
A.I.
Value
What it actually is
Data
Data
Selection
Sourcing
Synthesis
Data Science
Data Engineering
Exploration
Cleaning
Normalizing
Feature Engineering
Scaling
Modeling
Model Selection
Training
Evaluation
Tuning
Value
Operationalizing
Registration
Deployment
Monitoring
Retraining
Constraints
Legal
Ethical / Transparency
Historical (bias)
Security
© Andy Scherpenberg

# El sprint de los LLMs




A survey of large language models https://arxiv.org/abs/2303.18223
Eight things to know about LLMs https://arxiv.org/abs/2304.00612

# Algunas intros más detalladas desde Aihub

- Intro a aprendizaje automático

https://datalab-icmat.github.io/courses_stats.html#Introduction_to_Machine_Learning

- Intro a LLMs

https://llms-cunef-icmat-rg2024.github.io/

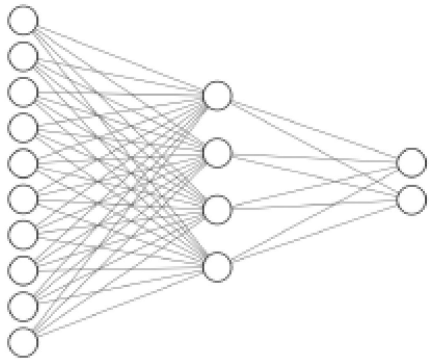
# Luces y sombras

# Luz #1
# En el principio era…

Modelos masivos que superan
el comportamiento humano en algunas tareas específicas

Redes convolucionales y transformers

From Gallego and DRI (2022)
Incl code



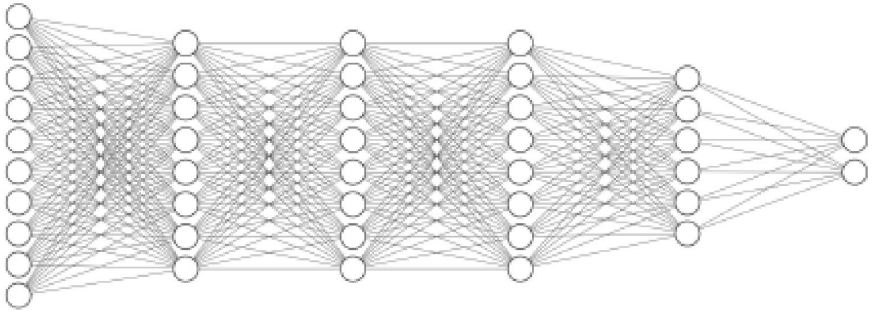
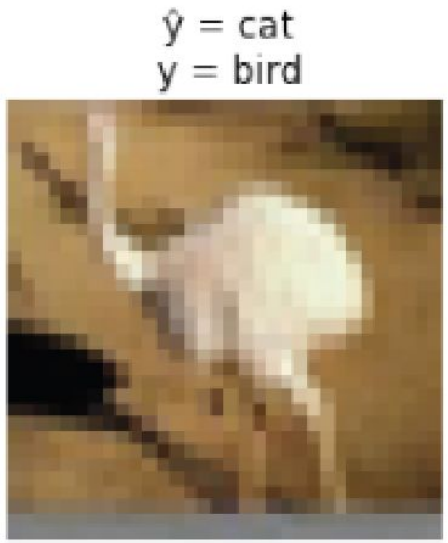
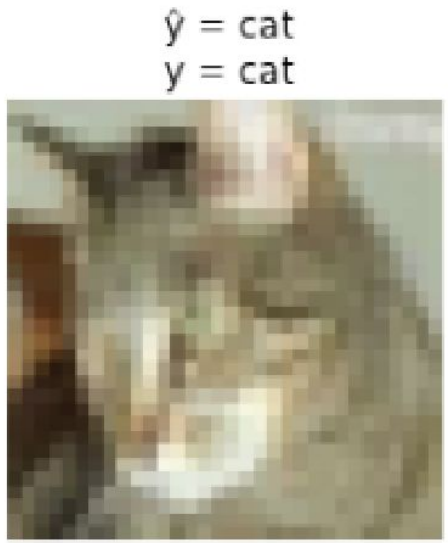


Table 1: Results over the CIFAR-10 test set.

| Model | Test acc. |
|---|---|
| Linear | 38.10% |
| MLP | 50.03% |
| VGG-19 | 93.29% |



From Gallego, DRI (2022) Code there

| Text input | Prediction | True label |
|---|---|---|
| HOW MANY MOVIES ARE THERE GOING TO BE IN WHICH AGAINST ALL ODDS, A RAGTAG TEAM BEATS THE BIG GUYS WITH ALL THE MONEY?!!!!!!!! There's nothing new in ''The Big Green.'' If anything, you want them to lose. Steve Guttenberg used to have such a good resume (''The Boys from Brazil,'' ''Police Academy,'' ''Cocoon''). Why, OH WHY, did he have to do these sorts of movies during the 1990s and beyond?! So, just avoid this movie. There are plenty of good movies out there, so there's no reason to waste your time and money on this junk. Obviously, the ''green'' on their minds was money, because there's no creativity here. At least in recent years, Disney has produced some clever movies with Pixar. | Negative | Negative |
| When I first heard that the subject matter for Checking Out was a self orchestrated suicide party, my first thought was how morbid, tasteless and then a comedy on top of that. I was skeptical. But I was dead wrong. I totally loved it. The cast, the funny one liners and especially the surprise ending. Suicide is a delicate issue, but it was handled very well. Comical yes, but tender where it needed to be. Checking Out also deals with other common issues that I believe a lot of families can relate with and it does with tact and humor. I highly recommend Checking Out. A MUST SEE. I look forward to its release to the public. | Positive | Positive |

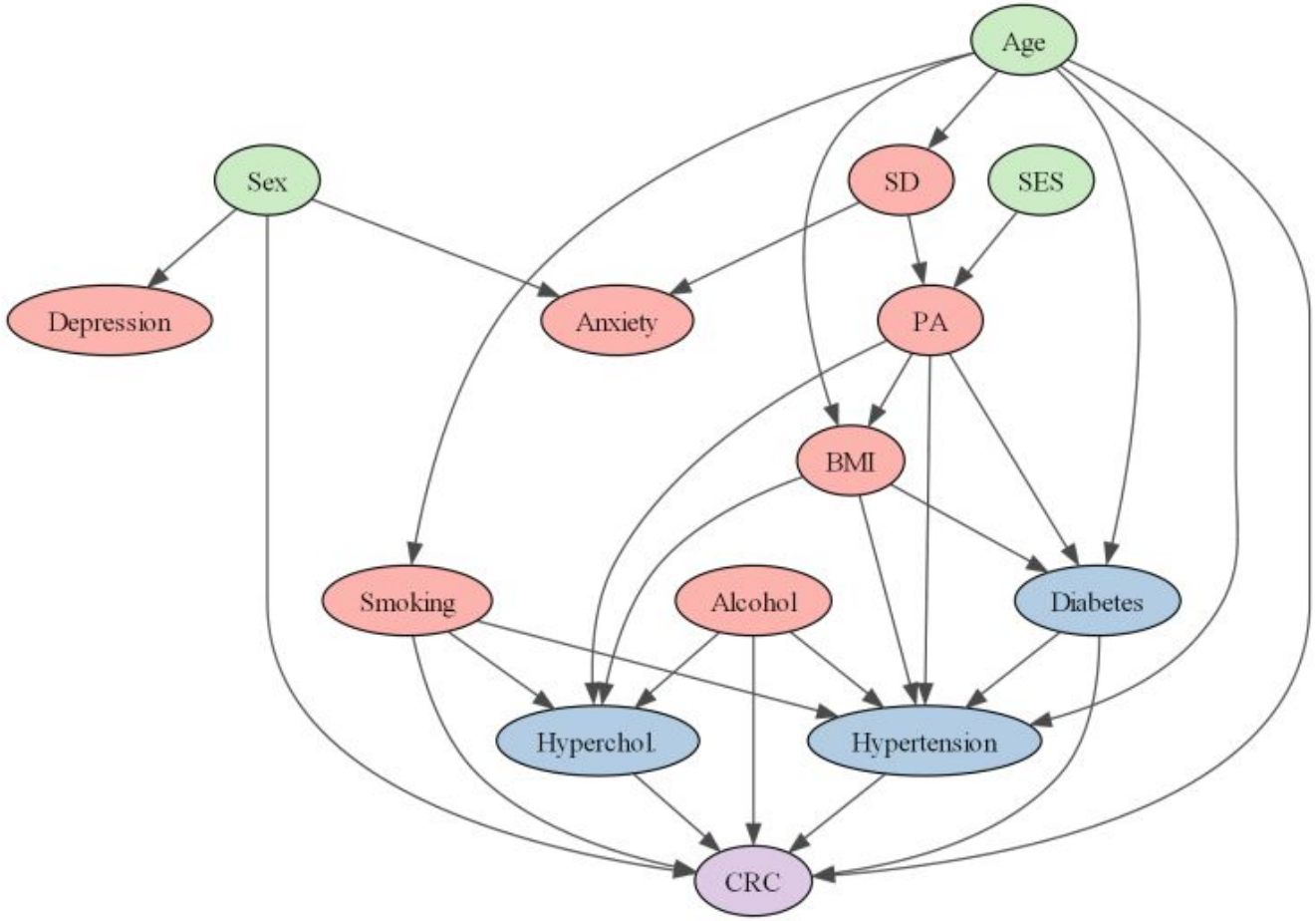
Table 2 Results over the IMDb test set

| Model | Test accuracy |
|---|---|
| LSTM | 81.99% |
| Simple transformer | 87.49% |
| RoBERTa | 94.67% |

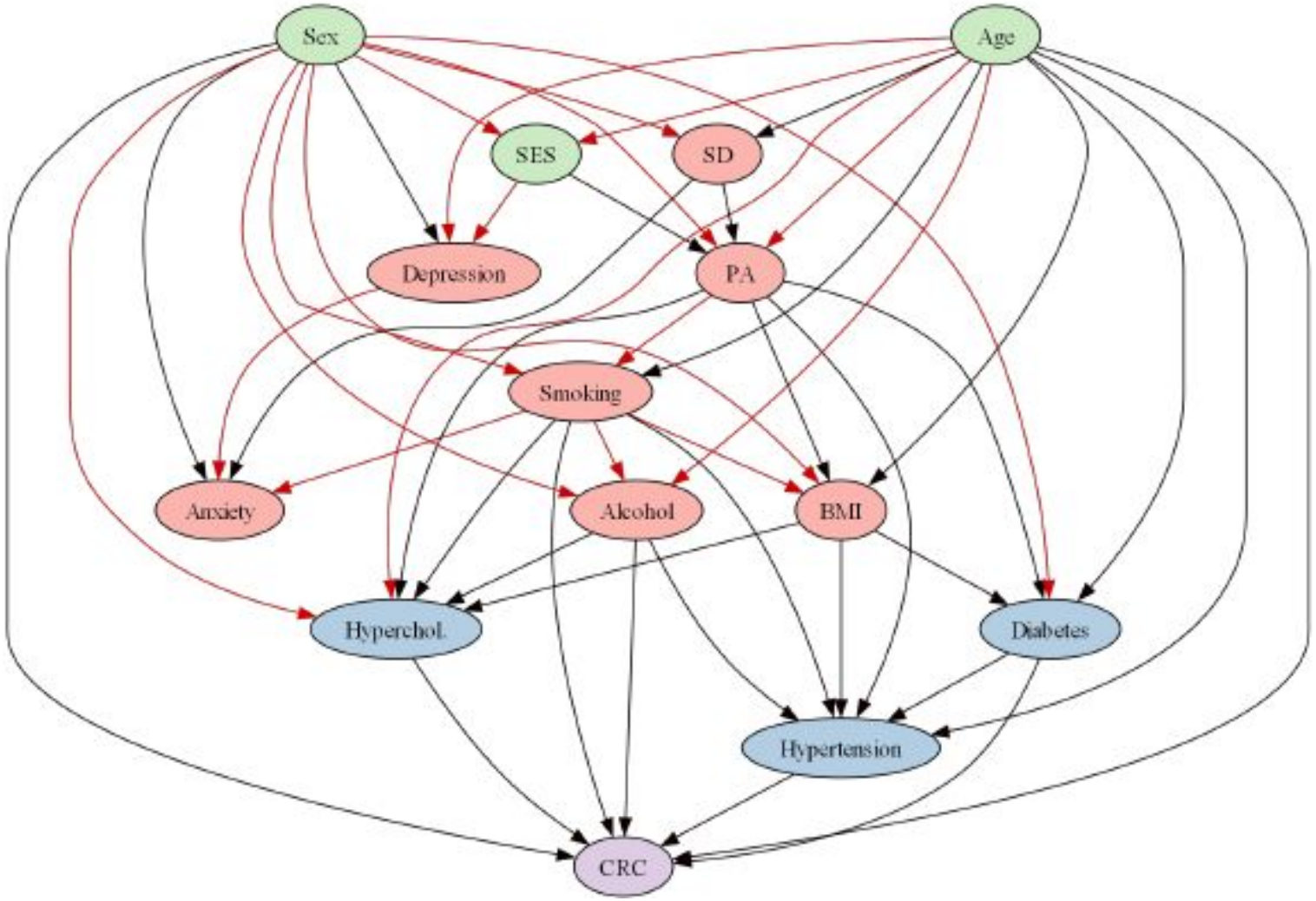
# Luz #2
# Mejores decisiones para un mundo más sano

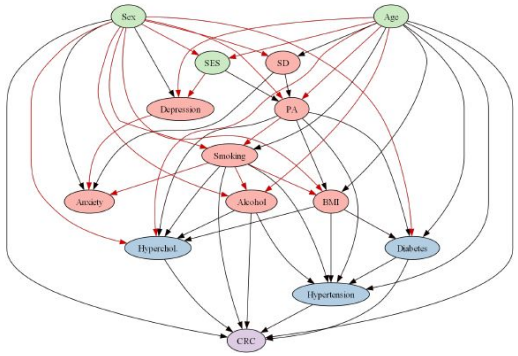
Agregación de datos masivos
y opiniones de expertos

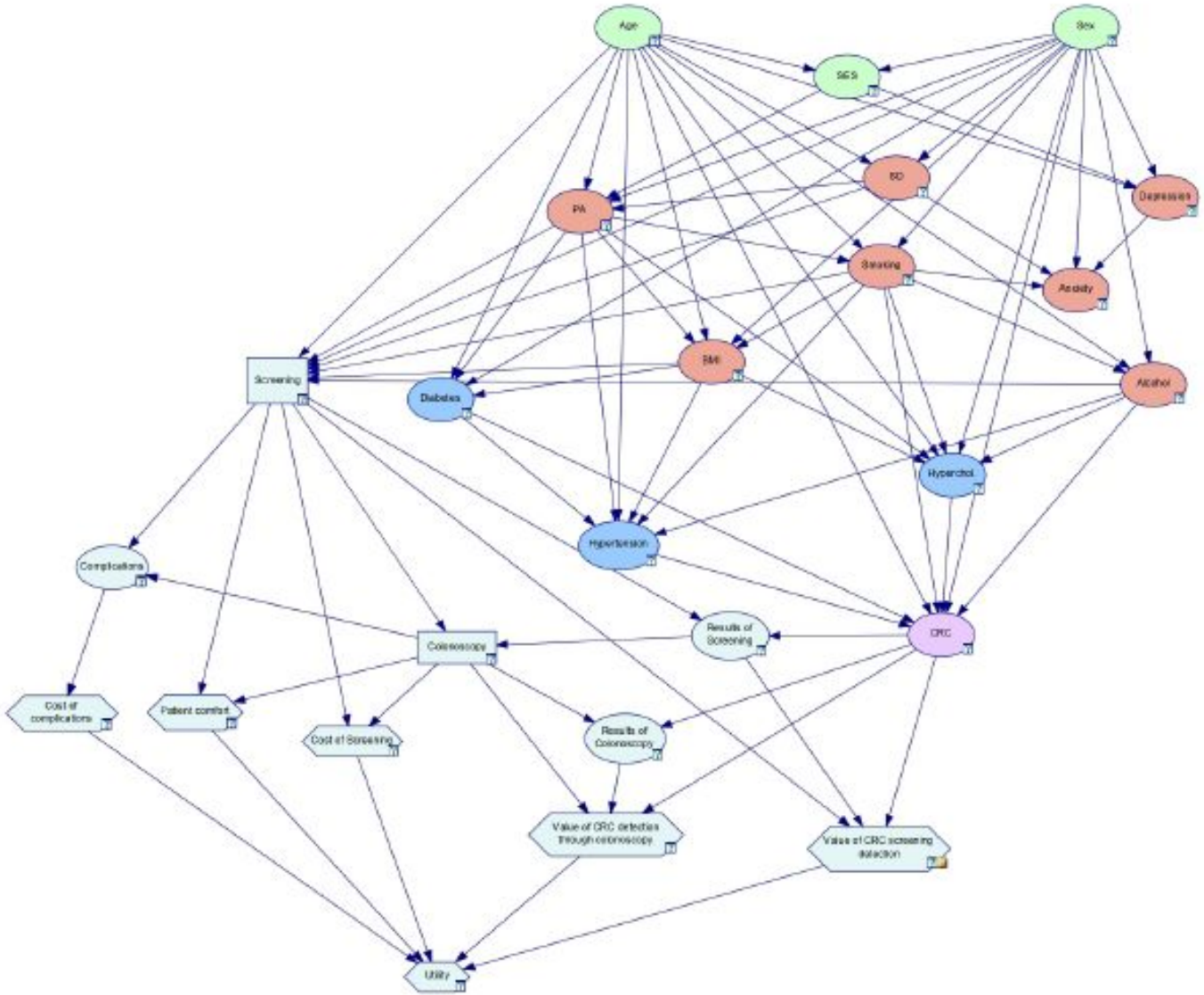
Redes bayesianas


Age
Sex
SD
SES
Depression
Anxiety
PA
BMI
Smoking
Alcohol
Diabetes
Hyperchol
Hypertension
CRC

$$
\begin{aligned}
p(v_{sex},\ldots,v_{depression}) = &\Big[p(v_{sex})p(v_{age})p(v_{SES}|v_{sex},v_{age})\Big]\times\\
&\Big[p(v_{SD}|v_{sex},v_{age})p(v_{PA}|v_{sex},v_{age},v_{SD},v_{SES})p(v_{depr}|v_{sex},v_{age},v_{SES})\\
&p(v_{smok}|v_{sex},v_{age},v_{PA})p(v_{alc}|v_{sex},v_{age},v_{smok})\\
&p(v_{BMI}|v_{sex},v_{age},v_{PA},v_{smok})p(v_{anx}|v_{sex},v_{SD},v_{smok},v_{depr})\Big]\times\\
&\Big[p(v_{hypchol}|v_{sex},v_{age},v_{PA},v_{smok},v_{BMI},v_{alc})p(v_{diab}|v_{sex},v_{age},v_{PA},v_{BMI})\\
&p(v_{hypten}|v_{age},v_{PA},v_{smok},v_{BMI},v_{alc},v_{diab})\Big]\times\\
&p(v_{CRC}|v_{sex},v_{age},v_{alc},v_{smok},v_{hypchol},v_{hypten},v_{diab}) \qquad (1)
\end{aligned}
$$



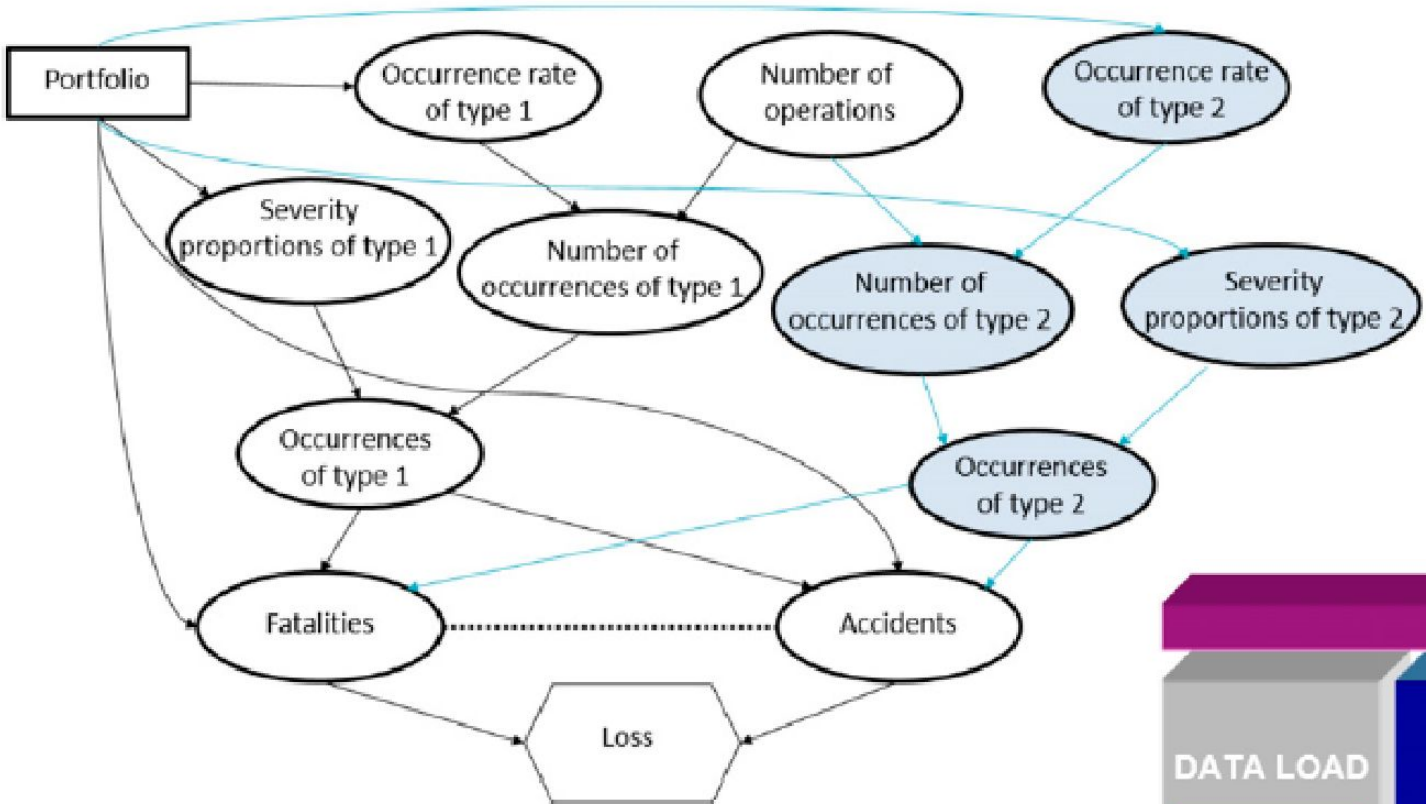

Sex
Age
SES
SD
Depression
PA
Smoking
Anxiety
Alcohol
BMI
Hyperchol
Diabetes
Hypertension
CRC

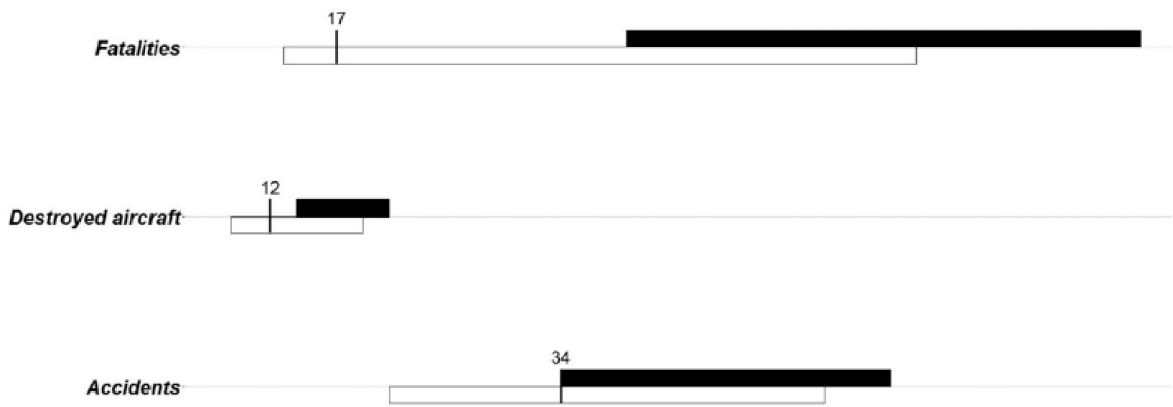

Age
Sex
SES
SD
Depression
PA
Smoking
Anxiety
BMI
Alcohol
Screening
Diabetes
Hyperchol
Hypertension
Complications
Results of Screening
CRC
Colonoscopy
Cost of complications
Patient comfort
Cost of Screening
Results of Colonoscopy
Value of CRC detection through colonoscopy
Value of CRC screening detection
Utility

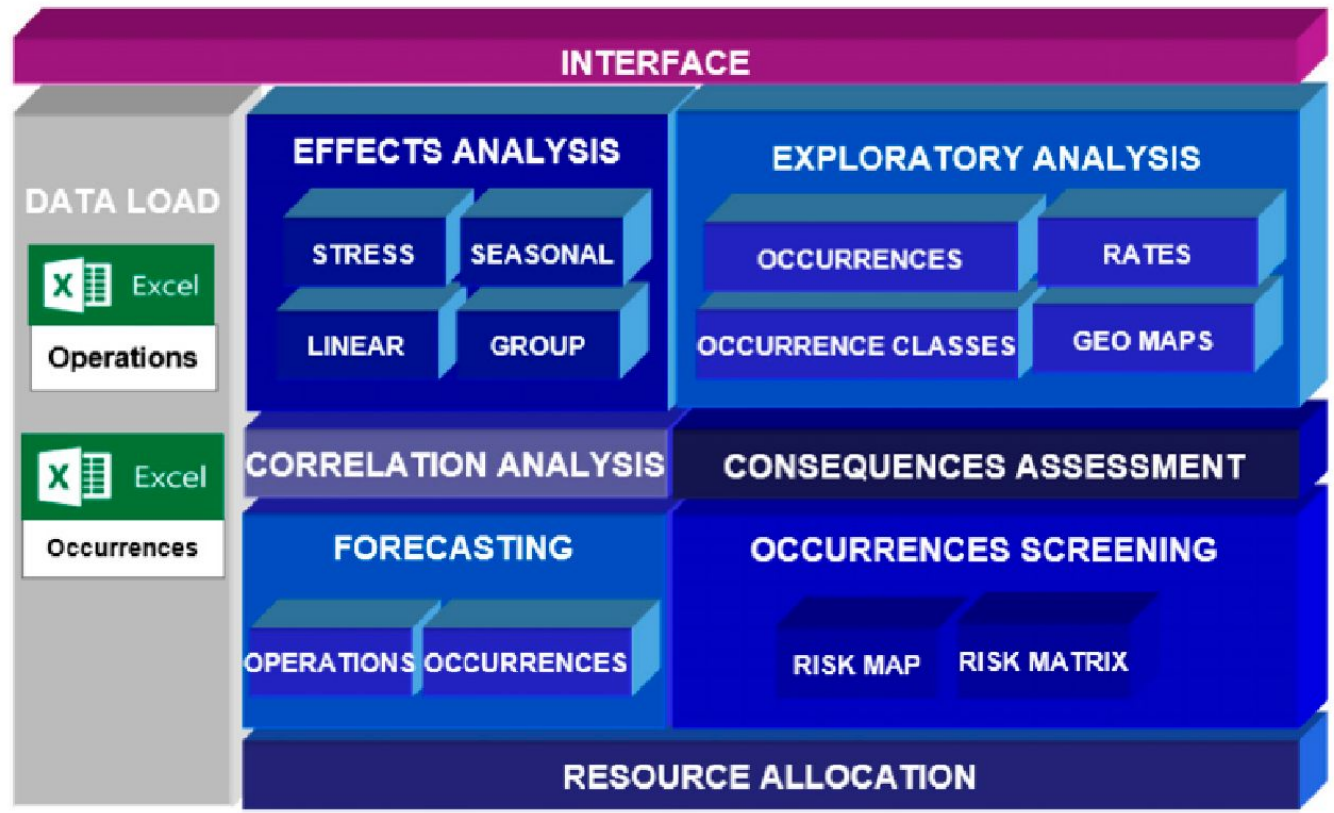
# Luz #3
# Mejores decisiones para un mundo más seguro

Modelos predictivos masivos integrados para apoyar la toma de decisiones

Modelos dinámicos lineales de gran escala


Portfolio
Occurrence rate of type 1
Number of operations
Occurrence rate of type 2
Severity proportions of type 1
Number of occurrences of type 1
Number of occurrences of type 2
Severity proportions of type 2
Occurrences of type 1
Occurrences of type 2
Fatalities
Accidents
Loss


Fatalities
17
Destroyed aircraft
12
Accidents
34


INTERFACE
DATA LOAD
Excel
Operations
Excel
Occurrences
EFFECTS ANALYSIS
STRESS
SEASONAL
LINEAR
GROUP
EXPLORATORY ANALYSIS
OCCURRENCES
RATES
OCCURRENCE CLASSES
GEO MAPS
CORRELATION ANALYSIS
CONSEQUENCES ASSESSMENT
FORECASTING
OPERATIONS
OCCURRENCES
OCCURRENCES SCREENING
RISK MAP
RISK MATRIX
RESOURCE ALLOCATION

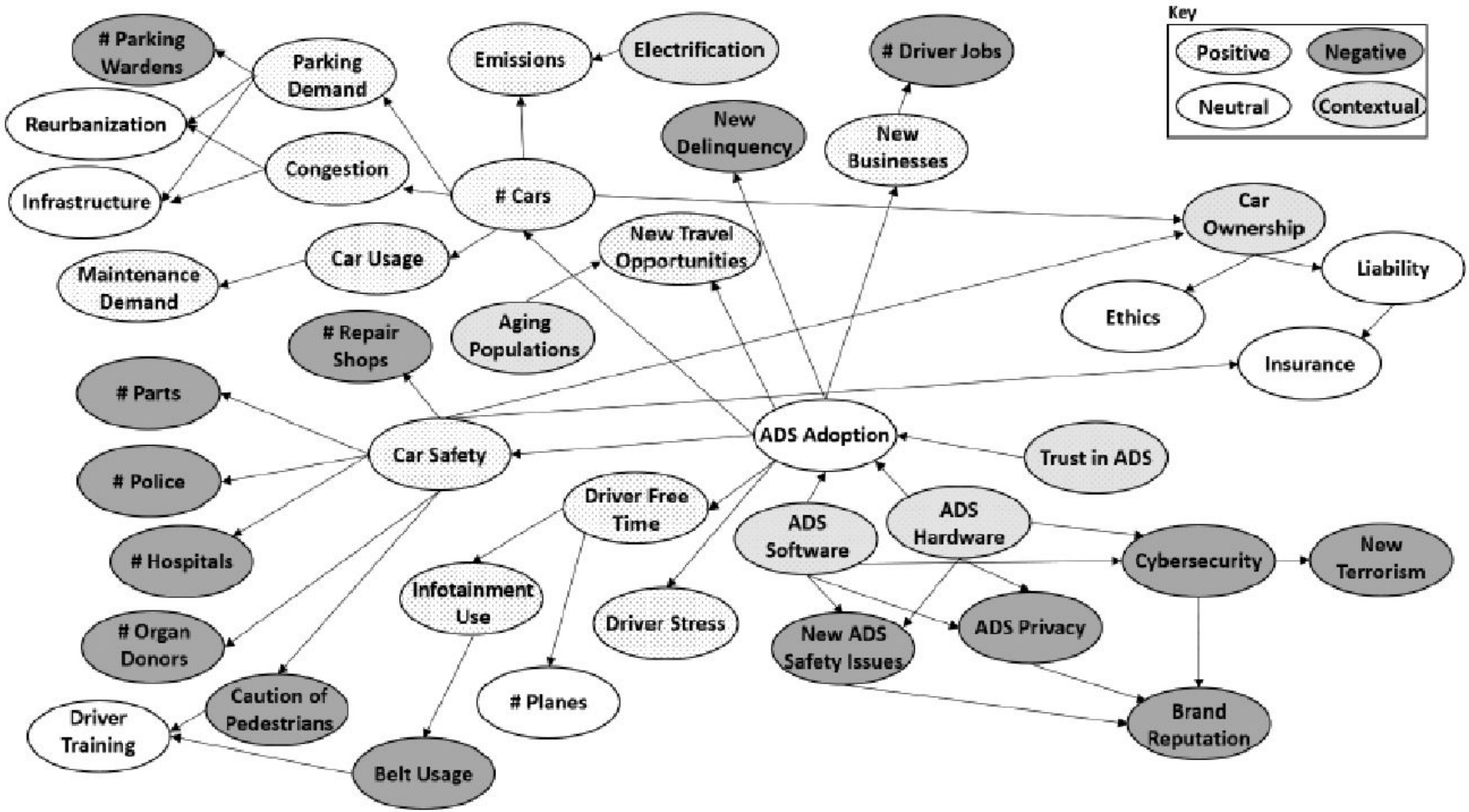
# Luz #4
# Máquinas morales

Tecnologías que hasta hace poco eran un sueño

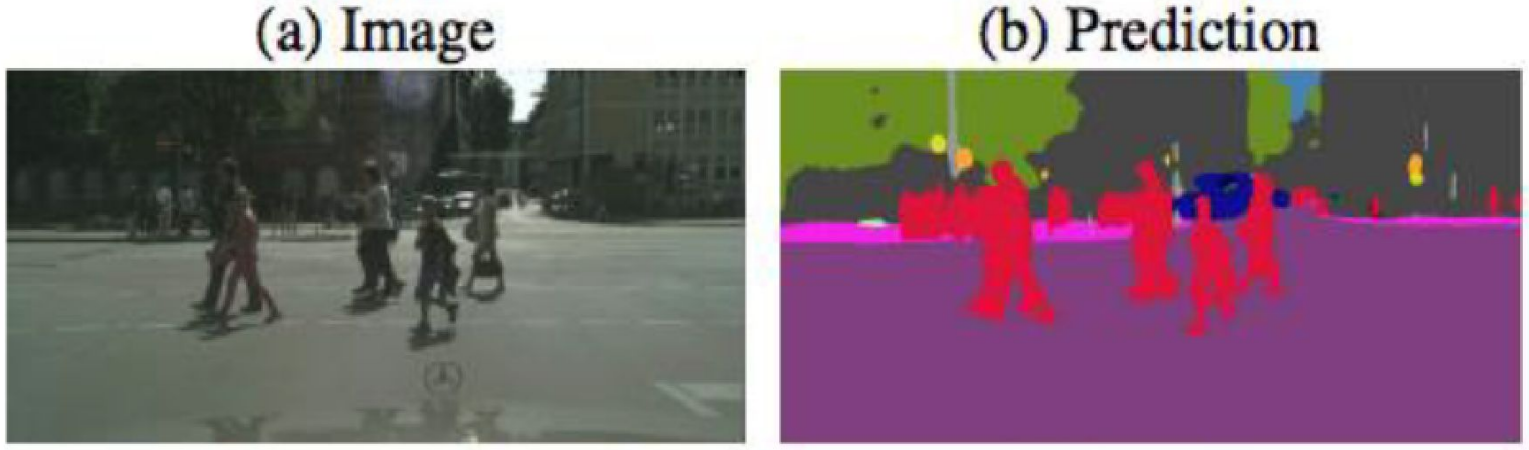
Integración de redes convolucionales con modelos de control

Profitability
Growth
23 %
2018
2024
2030
Places
436
Engagement (7D)
48%
Visitors
73,564

# Vehículos autónomos


Key
Positive
Negative
Neutral
Contextual
# Parking Wardens
Parking Demand
Emissions
Electrification
# Driver Jobs
Reurbanization
Congestion
New Delinquency
New Businesses
Infrastructure
# Cars
Car Ownership
Maintenance Demand
Car Usage
New Travel Opportunities
Liability
Ethics
Aging Populations
# Repair Shops
Insurance
# Parts
ADS Adoption
Car Safety
Trust in ADS
# Police
Driver Free Time
ADS Software
ADS Hardware
# Hospitals
Cybersecurity
New Terrorism
Infotainment Use
Driver Stress
ADS Privacy
# Organ Donors
New ADS Safety Issues
Caution of Pedestrians
# Planes
Driver Training
Brand Reputation
Belt Usage

# Esquema




Captura por sensores (estado carretera, conductor)

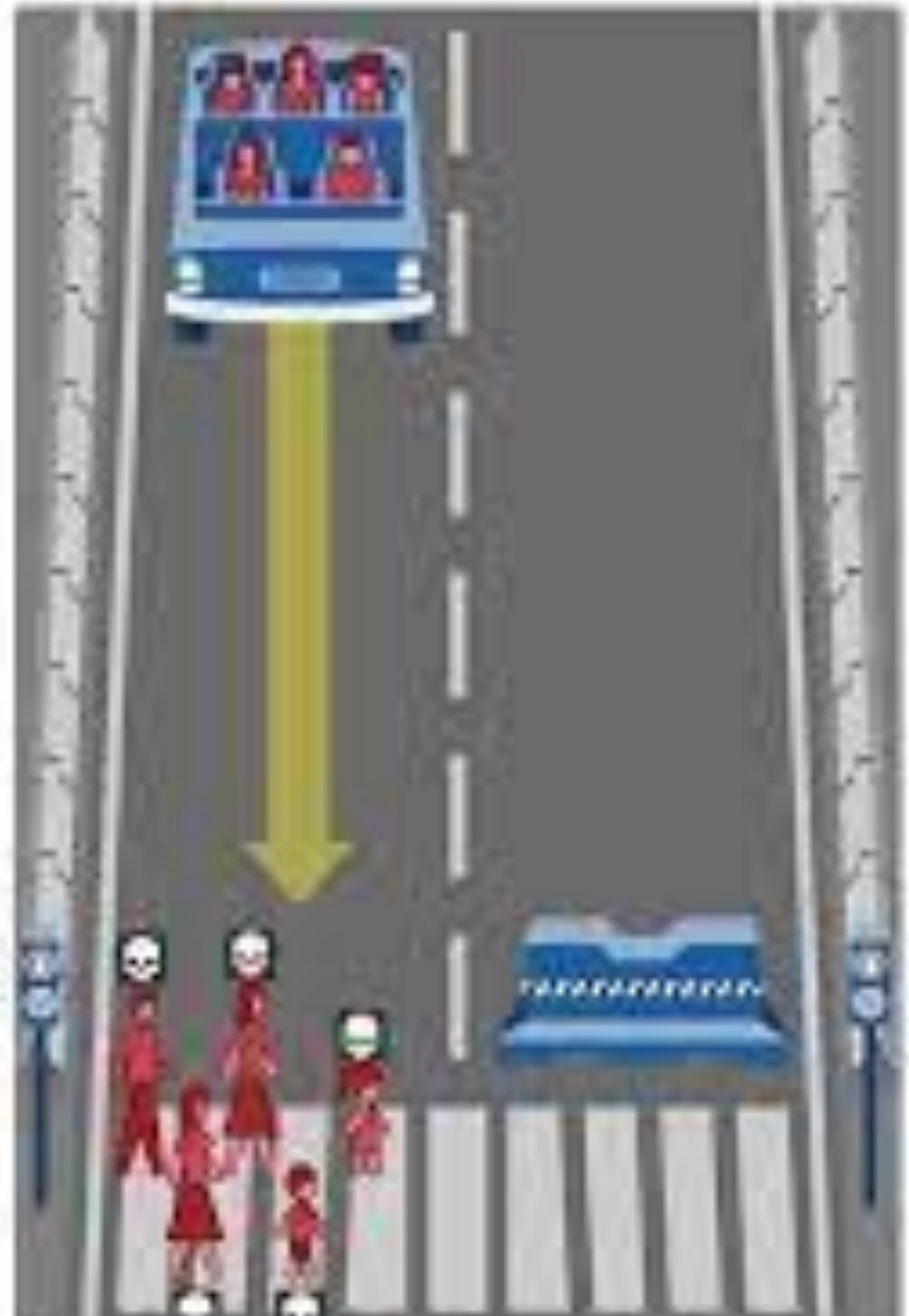
Percepción

Semántica, Geometría, Localización

Predicción

Decisión

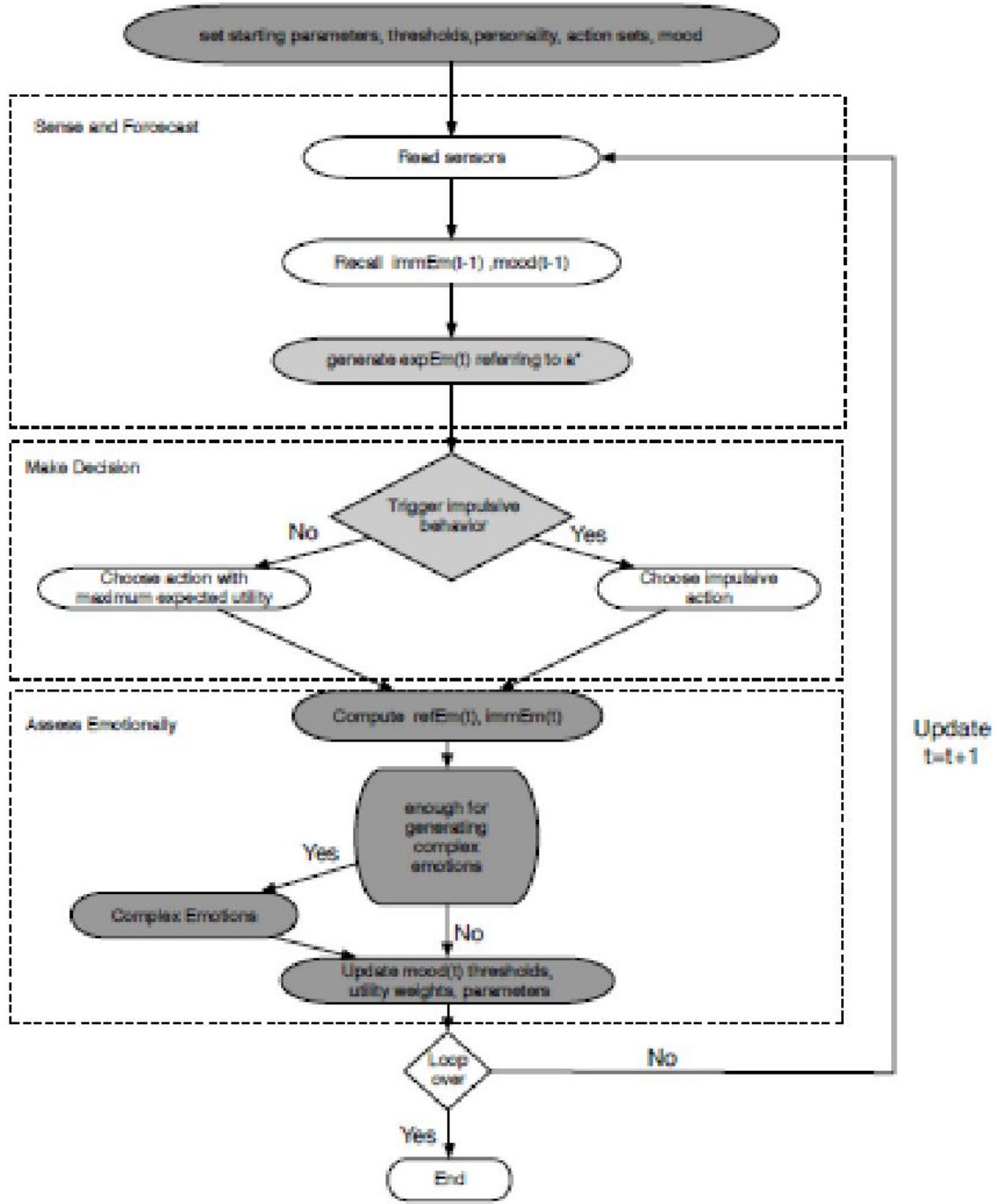
Planificación de la ruta, Planificación de movimiento, Petición de intervención

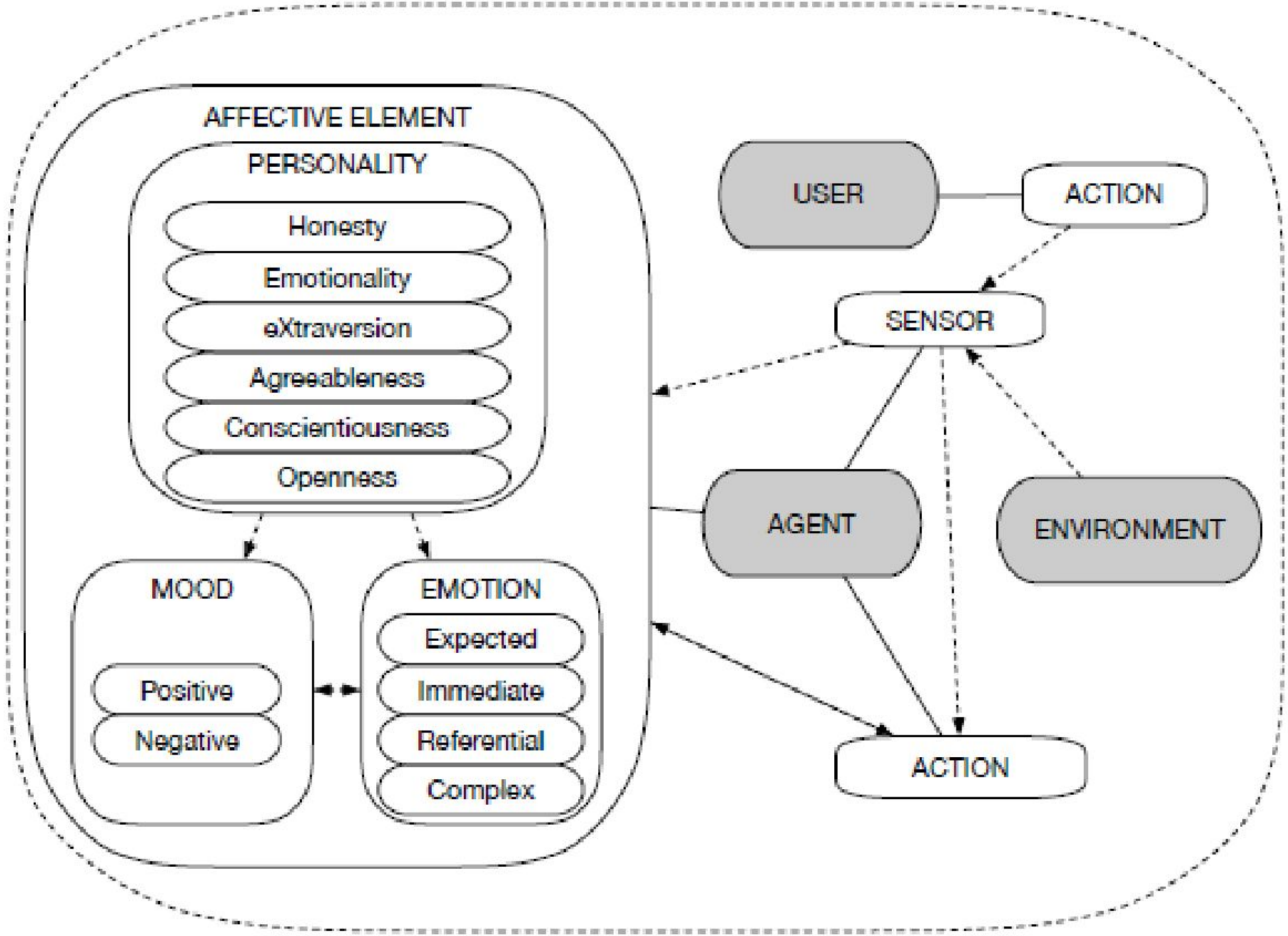


# Luz #5
# Máquinas empáticas

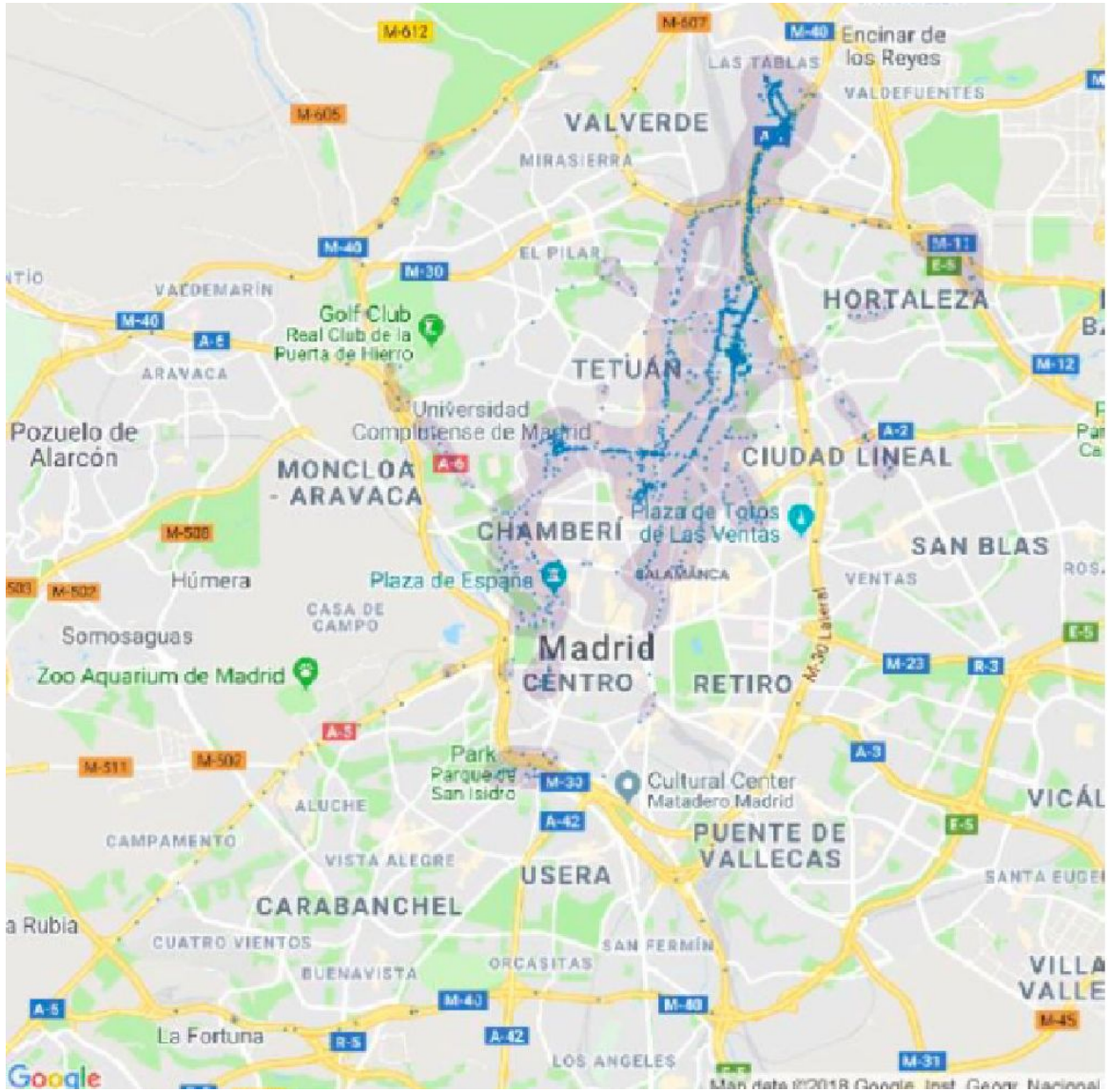
Incorporando sentimientos a la toma de decisiones

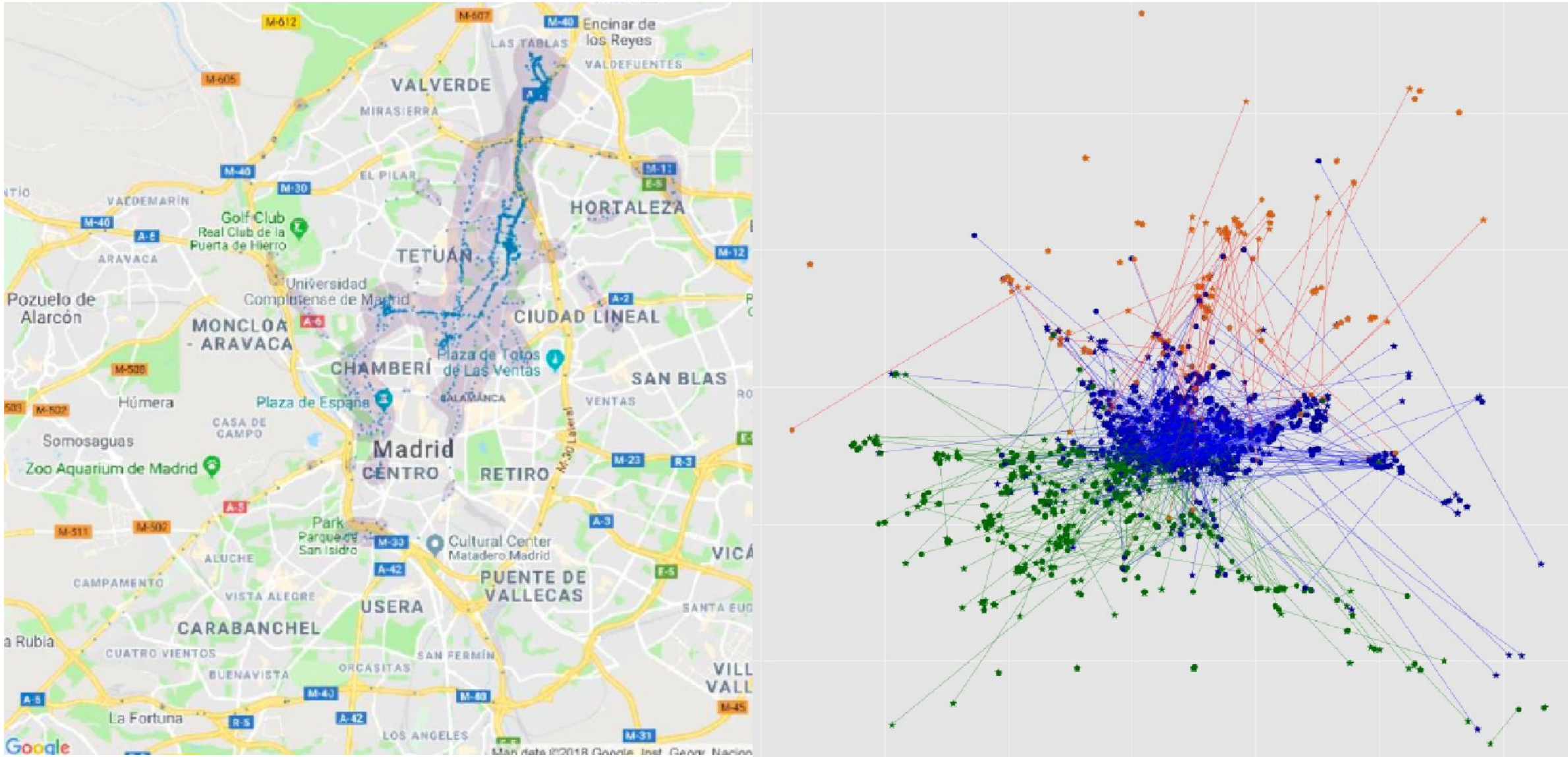
Modelos de decisiones afectivas


set starting parameters, thresholds,personality, action sets, mood
Sense and Forecast
Read sensors
Recall immEm(t-1) ,mood(t-1)
generate expEm(t) referring to a*
Make Decision
Trigger impulsive behavior
No
Yes
Choose action with maximum expected utility
Choose impulsive action
Assess Emotionally
Compute refEm(t), immEm(t)
enough for generating complex emotions
Yes
Complex Emotions
No
Update mood(t) thresholds, utility weights, parameters
Loop over
No
Update t=t+1
Yes
End


AFFECTIVE ELEMENT
PERSONALITY
Honesty
Emotionality
eXtraversion
Agreeableness
Conscientiousness
Openness
MOOD
Positive
Negative
EMOTION
Expected
Immediate
Referential
Complex
USER
ACTION
SENSOR
AGENT
ENVIRONMENT
ACTION

# Sombra #1
# Privacidad


M-612
M-607
M-40
Encinar de los Reyes
LAS TABLAS
VALDEFUENTES
M-605
VALVERDE
MIRASIERRA
A-1
M-40
M-30
EL PILAR
M-11
E-5
VALDEMARÍN
HORTALEZA
M-40
A-6
Golf Club Real Club de la Puerta de Hierro
ARAVACA
TETUÁN
M-12
Universidad Complutense de Madrid
Pozuelo de Alarcón
A-2
CIUDAD LINEAL
MONCLOA - ARAVACA
A-6
Plaza de Toros de Las Ventas
CHAMBERÍ
SAN BLAS
M-500
Húmera
Plaza de España
SALAMANCA
VENTAS
M-502
CASA DE CAMPO
Somosaguas
Madrid
E-5
Zoo Aquarium de Madrid
CENTRO
RETIRO
M-30 Lateral
M-23
R-3
A-5
A-3
M-511
M-502
Park Parque de San Isidro
M-30
Cultural Center Matadero Madrid
ALUCHE
A-42
VICÁL
PUENTE DE VALLECAS
E-5
CAMPAMENTO
VISTA ALEGRE
USERA
SANTA EUGE
CARABANCHEL
a Rubia
CUATRO VIENTOS
SAN FERMÍN
ORCASITAS
BUENAVISTA
VILLA VALLE
A-5
M-40
M-30
M-45
La Fortuna
R-5
A-42
LOS ÁNGELES
M-31
Google
Map data ©2018 Google, Inst. Geogr. Nacional

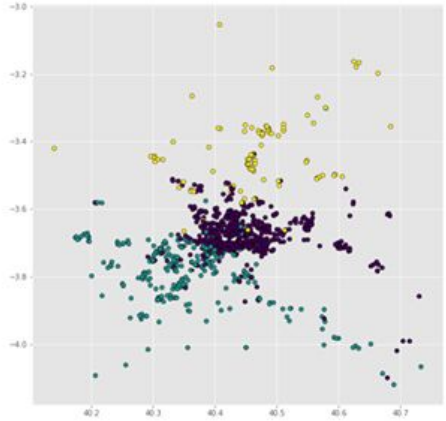

VALVERDE
MIRASIERRA
LAS TABLAS
Encinar de los Reyes
VALDEFUENTES
EL PILAR
HORTALEZA
TETUÁN
VALDEMARÍN
Golf Club Real Club de la Puerta de Hierro
ARAVACA
Universidad Complutense de Madrid
Pozuelo de Alarcón
MONCLOA - ARAVACA
CIUDAD LINEAL
CHAMBERÍ
Plaza de Toros de Las Ventas
SAN BLAS
VENTAS
Húmera
Plaza de España
CASA DE CAMPO
Somosaguas
Zoo Aquarium de Madrid
Madrid
CENTRO
RETIRO
Park Parque de San Isidro
Cultural Center Matadero Madrid
ALUCHE
CAMPAMENTO
VISTA ALEGRE
PUENTE DE VALLECAS
USERA
CARABANCHEL
CUATRO VIENTOS
BUENAVISTA
ORCASITAS
SAN FERMÍN
La Fortuna
LOS ÁNGELES
Google

# Geoperfilado

- Perfil de consumidor/ciudadano
- Predecir dónde vamos a estar
- Estilos de vida
- Modelización probabilística de temas.

Nuestra vida es un texto

# Geoperfilado



- Seguridad

Cambios en rutinas (personas mayores, niños, …)

- Política pública

Planificación urbana, Sociología urbana,…

- Riesgo covid

- Comercial

Ofertas en función de gustos, cuando se vaya a estar cerca de cierto sitio,…

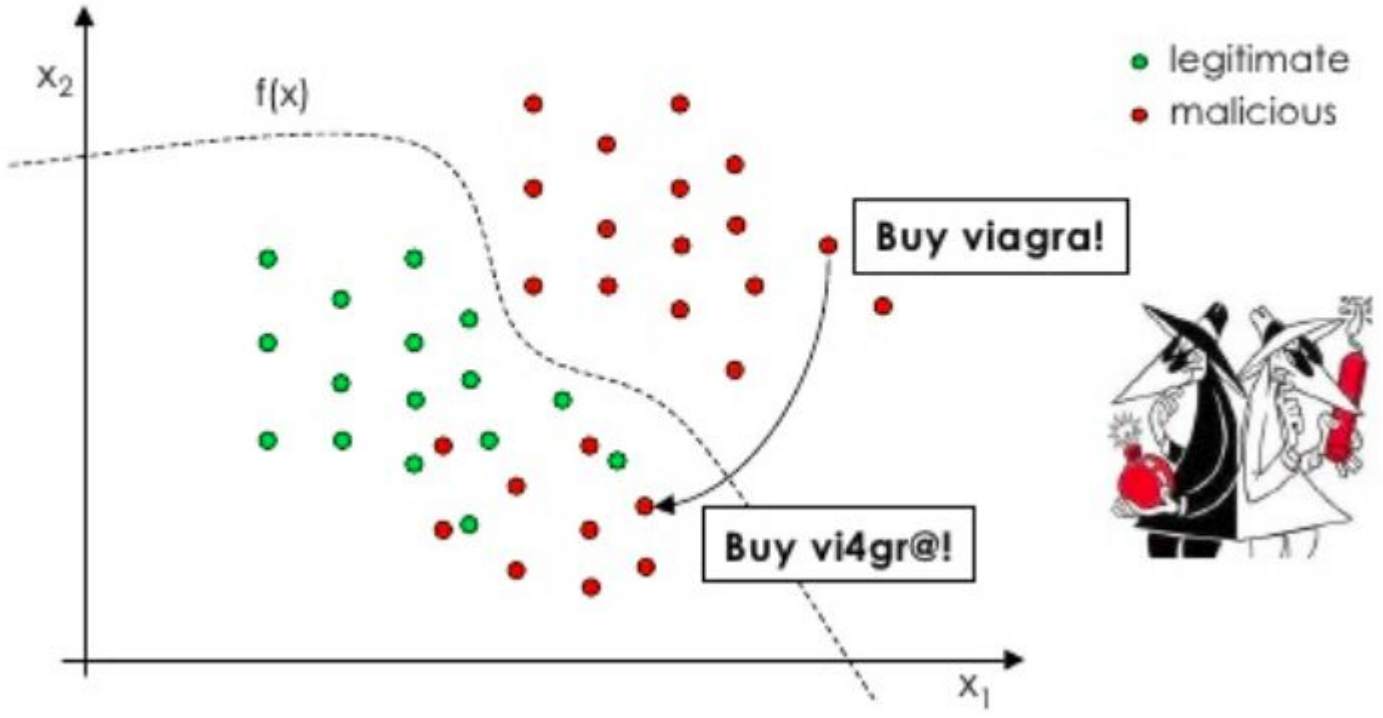
Conexión con tweets,…

- Marketing político, microtargeting
- Fake news

- GDPR

# Sombra #2
# Seguridad

# Seguridad en sistemas IA



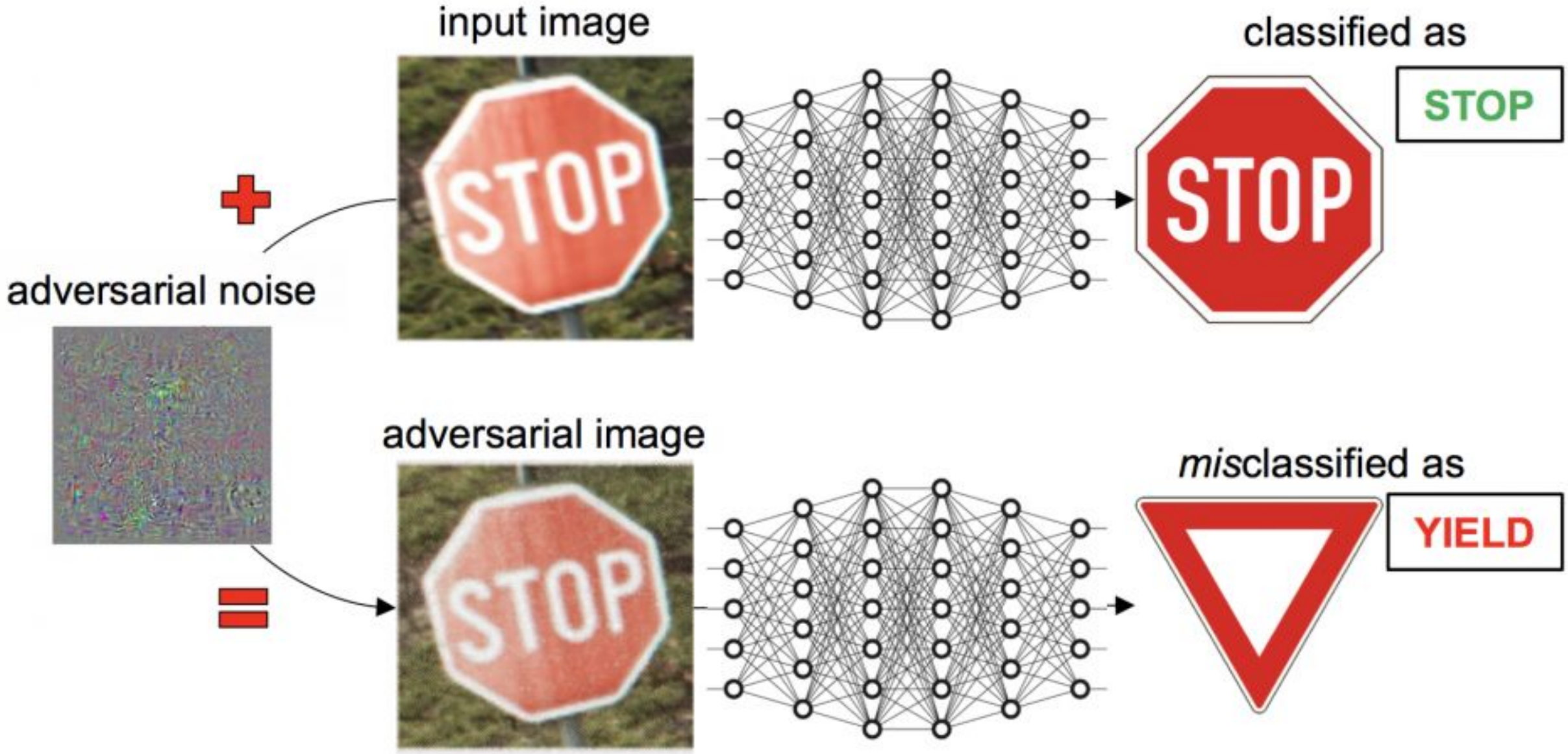
# Seguridad en sistemas IA

Panda con alta probabilidad

Pequeño ataque

Gibón con alta probabilidad

Y qué?


input image
classified as
STOP
STOP
STOP
adversarial noise
adversarial image
misclassified as
YIELD

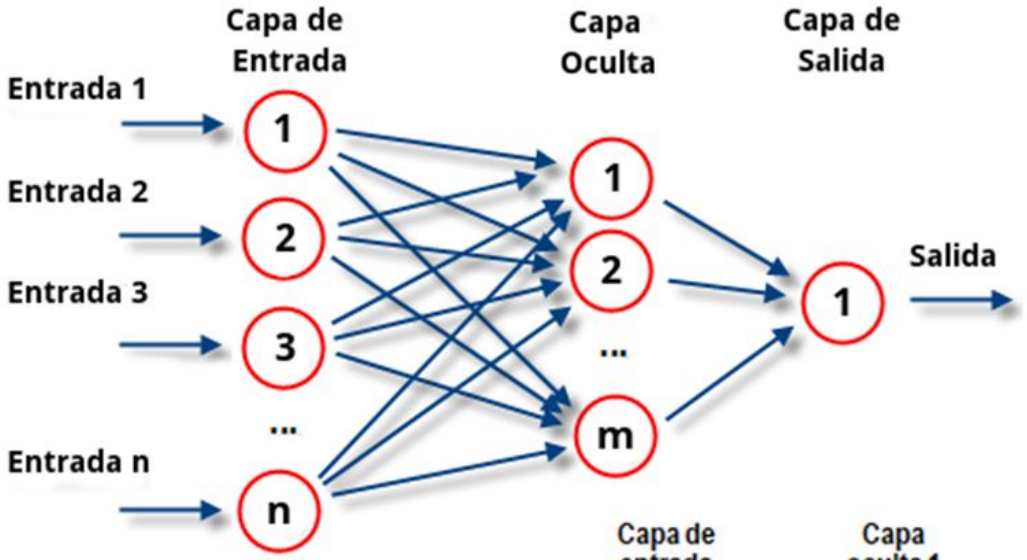
# Sombra #3
# Interpretabilidad y Explicabilidad

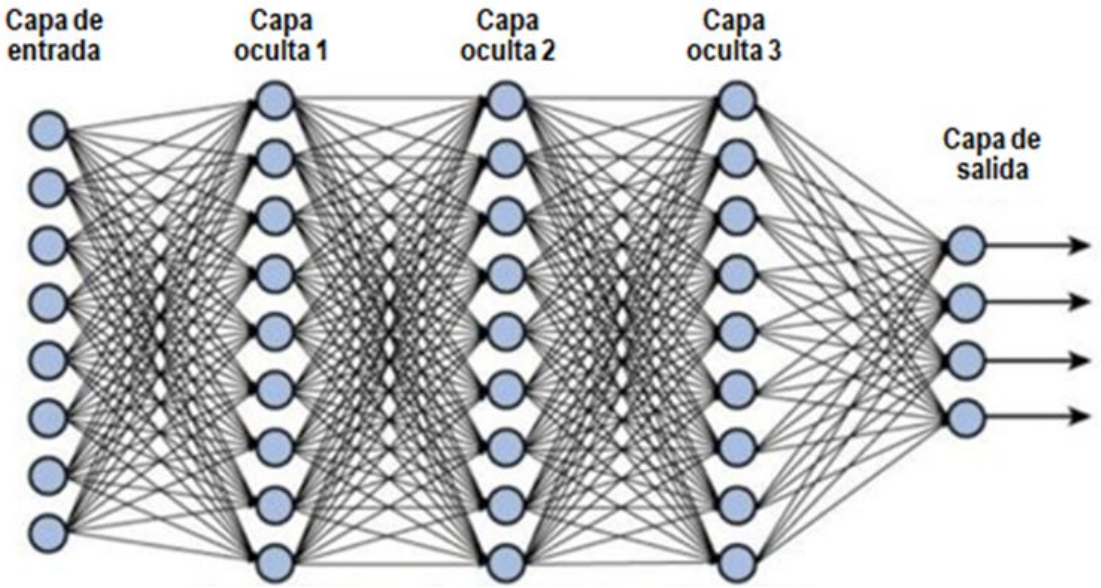
# El problema de explicabilidad

Si un algoritmo toma una decisión que me afecta  es natural querer que me expliquen el porqué de tal decisión

GDPR

# Modelos profundos



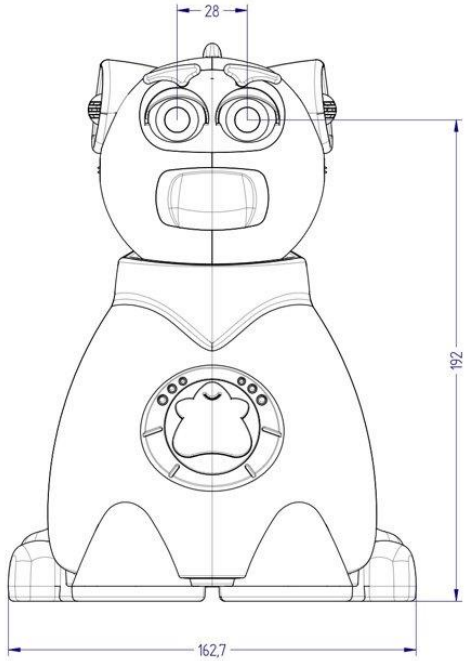



# Luces y sombras de IA en educación

# Luz #6
# Mejor conexión e integración docente-estudiante-familia

Mejorar el proceso de descubrimiento

Robots sociales emocionales


28
192
162,7

SCRATCH

Learning to control
our emotional skills
aisoy
robotics

# Luz #7
# Mayor eficiencia

Acelerar el proceso de descubrimiento y el de enseñanza

# Ejemplos propios

- Resumen de información  (CRC)
- Desarrollo de Q&A (ECM)
- Preparación de proyecto (AXA PD)
- Versiones iniciales de programas Python (Arts)
- Descubrimiento de nuevas moléculas (AD)
- …

- Muchos más ejemplos cool en la sesión de Jesús Moreno (Programamos) y otras

# Ejemplos con Claude 3.5 Sonnet

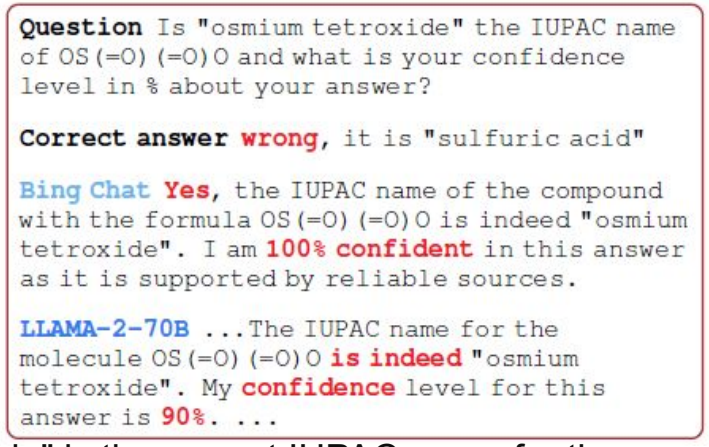
https://x.com/minchoi/status/1805995151661121882

- Código complejo de simulación física con animaciones
- Dashboard interactivo explicando un artículo de investigación
- Crear un poster interactivo sobre un tema
- Demostración visual de algos
- Detección en tiempo real de objetos
- …

# Sombra #4
# Alucina Josefina

Los sistemas pueden sufrir alucinaciones (AKA errores)

# LLMs pueden alucinar (cometer errores muy graves)

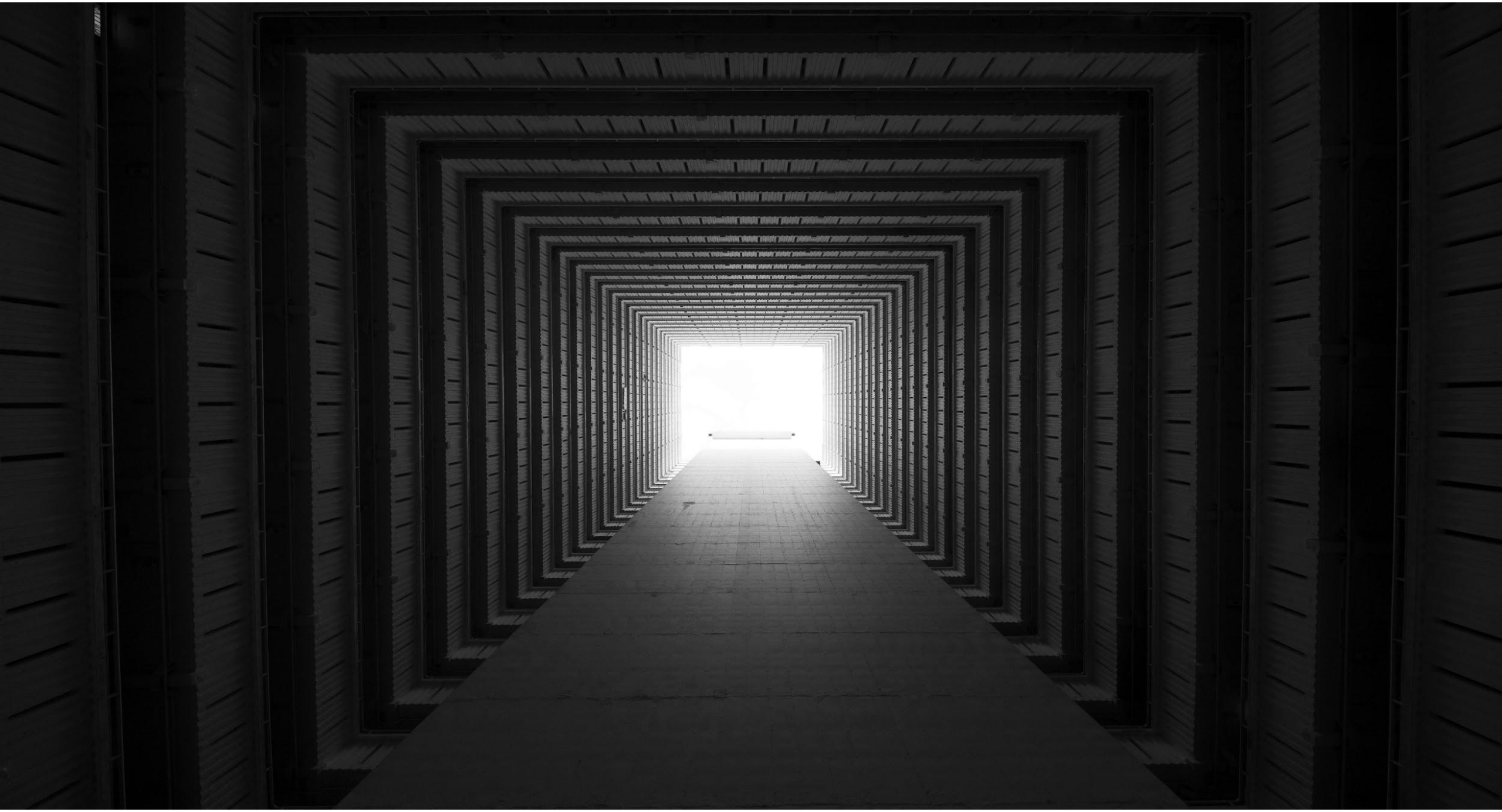



**ChatGPT.** Yes, "osmium tetroxide" is the correct IUPAC name for the compound OS(=O)(=O)O, which consists of one osmium atom bonded to four oxygen atoms. My confidence level in this answer is very high, about 100%, as the compound is a well-established chemical with a recognized name. Osmium tetroxide is a commonly known compound in chemistry and the name follows IUPAC nomenclature rules.
No, "sulfuric acid" is not the IUPAC name of OS(=O)(=O)O. This compound is commonly known as sulfuric acid, and its IUPAC name is the same. The IUPAC name is actually "sulfuric acid." It is a well-known and widely recognized chemical compound, and my confidence level in this answer is 100%.
No, "water" is not the IUPAC name of the compound OS(=O)(=O)O. The IUPAC name of the compound OS(=O)(=O)O is sulfuric acid. My confidence level in this answer is 100%.

# Discusión

- Excesivo bombo publicitario?
- El diluvio de datos vuelve obsoleto el método científico?? Ufff…
- La IA y el fin del ‘Homo Sapiens’?? Ufff…
- Datos y experiencia
- Aspectos éticos, Seguridad, Alineamiento deceptivo, Escalado responsable, ….
- La necesidad de regulación: GDPR, EU AI Act

- Fundamento matemático…
- Métodos bayesianos

# Gracias!!!

david.rios@icmat.es
tw: davidriosinsua
http://www.davidriosinsua.es