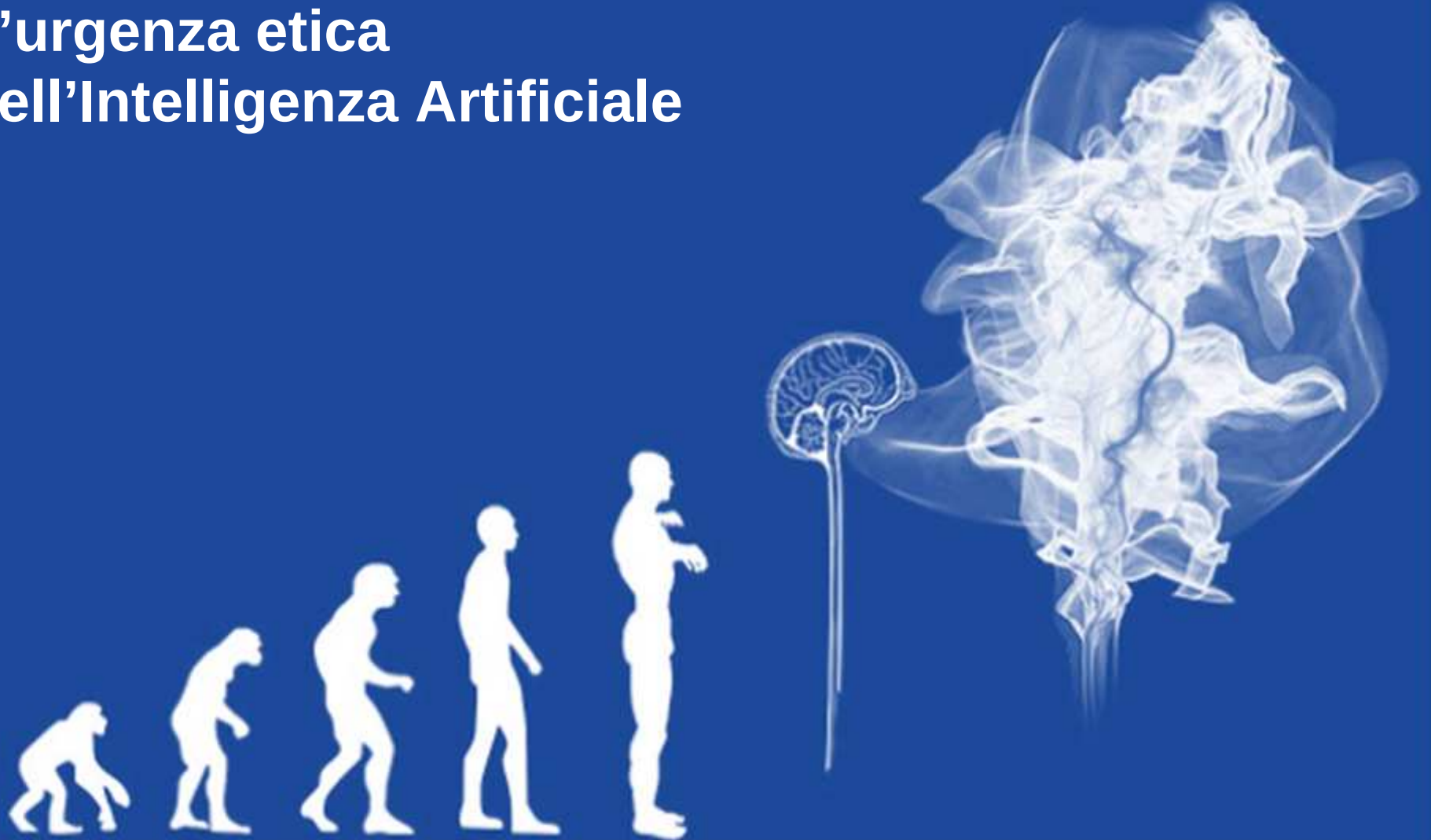


Etica Digitale nello sviluppo di soluzioni di Intelligenza Artificiale

L'urgenza etica dell'Intelligenza Artificiale



A. Ballarin

Società Generale d'Informatica SOGEI S.p.A.

1

sogei

Sede Legale Via M. Carucci n. 99 - 00143 Roma

Il ruolo dei dati nell'IA

Uno dei clichés della *Web economy*: **il dato è il nuovo petrolio**, e come ogni clichés contiene anche diverse verità ed implica conseguenze.

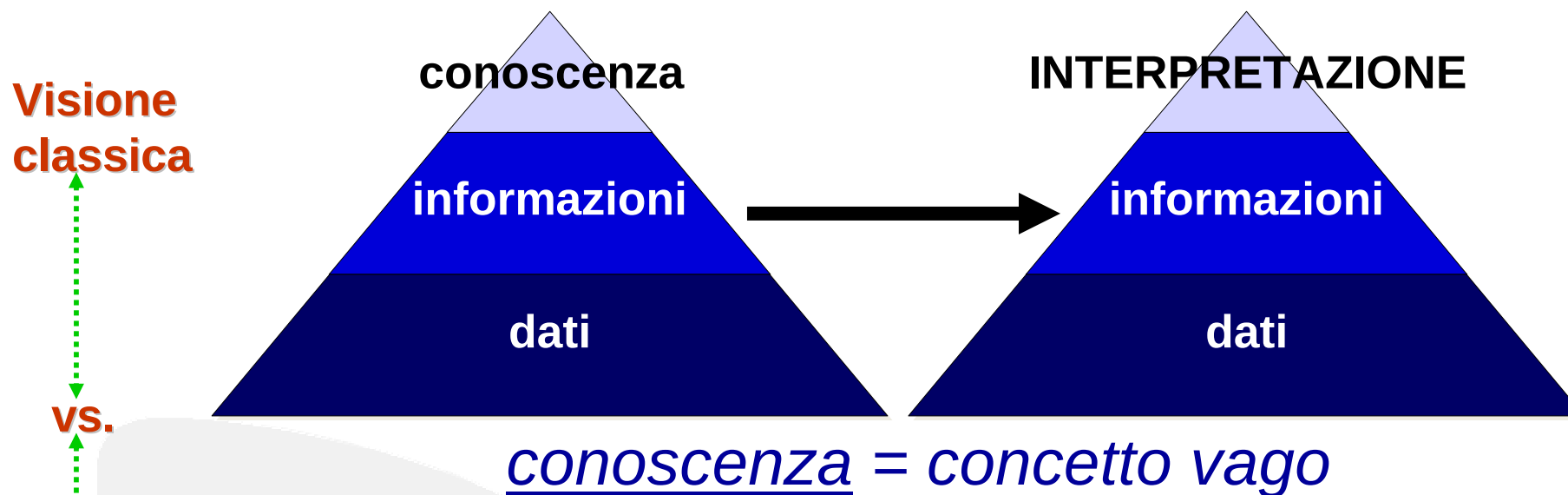
L'IA, nella tradizione di Turing ovvero nella tradizione del ***mechanical reasoning***, costruisce modelli di simulazione di comportamenti intelligenti. Mentre l'IA basata sull'attuale ***machine learning*** si fonda su algoritmi dove **la correlazione soppianta la causalità**.

Risultato:

Le scienze possono progredire senza modelli teorici coerenti, teorie unificate o un qualche tipo di spiegazione meccanicistica.



Dati: visione classica vs. moderno *machine learning*



F. J. Miller: *I = 0* (Information has no intrinsic meaning); *Information Research*, 8, 1, 2002.

Ruolo dei dati nel machine learning

I ragionamenti delle attuali **menti artificiali** sono **analisi algoritmiche complesse** basate sui dati che rinunciano al concetto di ***spiegabilità***.

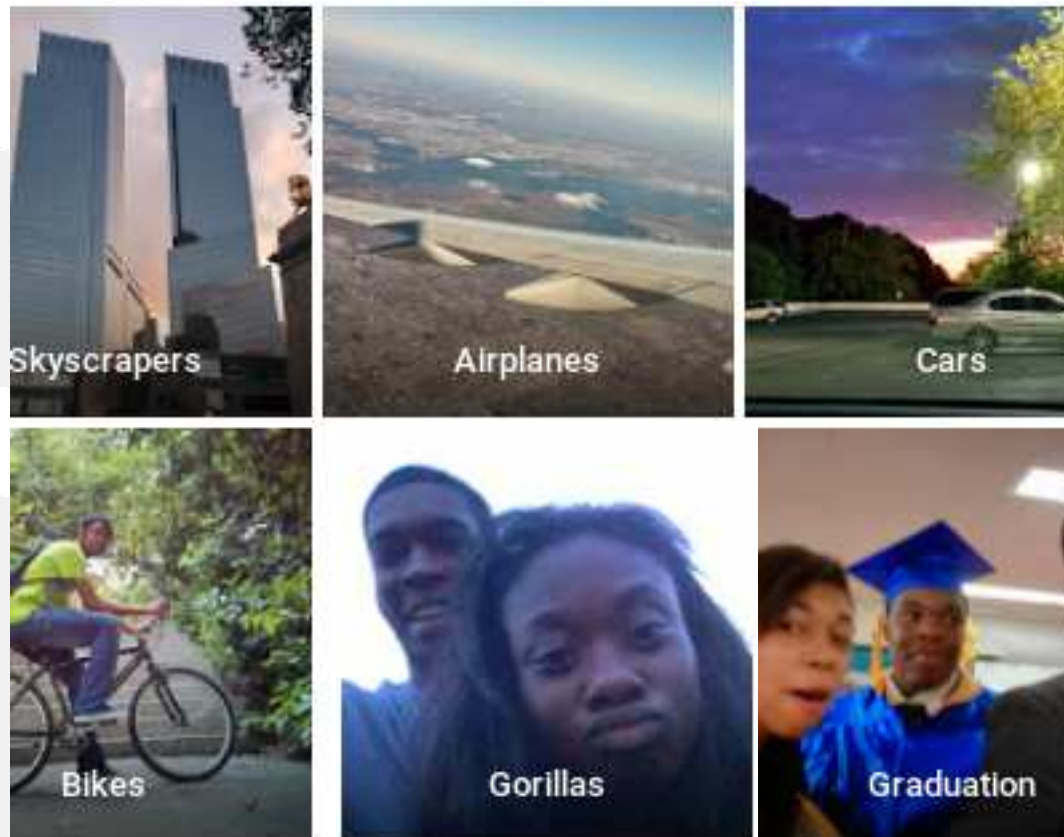
La **correlazione tra dati** viene usata per descrivere la realtà, pur senza avere alcuna teoria scientifica a supporto (es. diagnosi mediche, credit scoring, propensione alla spesa, ecc.).

L'IA non è esente da colossali cantonate

Rinunciando alla **spiegabilità**, quando l'IA prende cantonate sono, quasi sempre, **marchiane** e prive di un **perché logico**.

Google says sorry for racist auto-tag in photo app

The Guardian, 1° luglio 2015

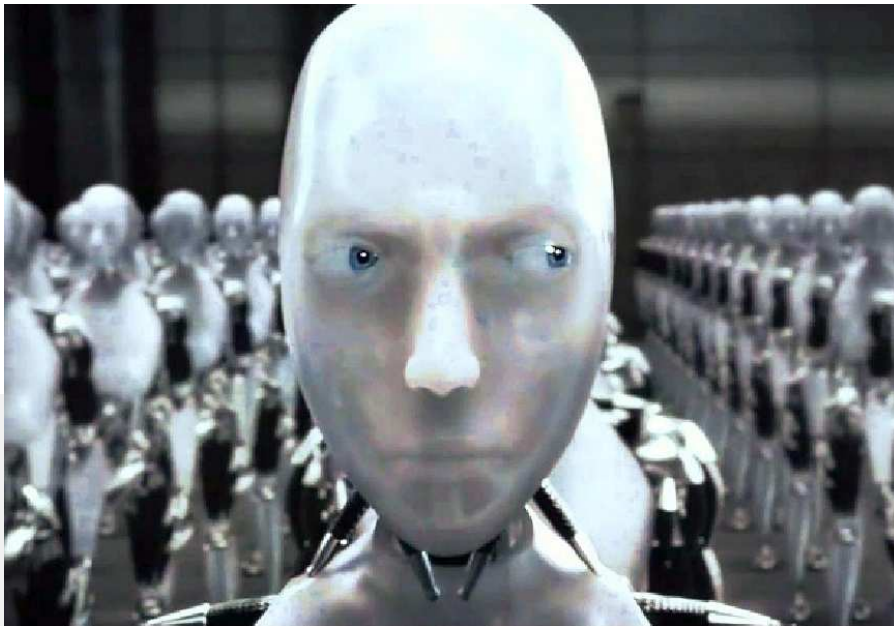


L'IA non è esente da colossali cantonate

L'algoritmo del male

Il Foglio, 16 settembre 2017

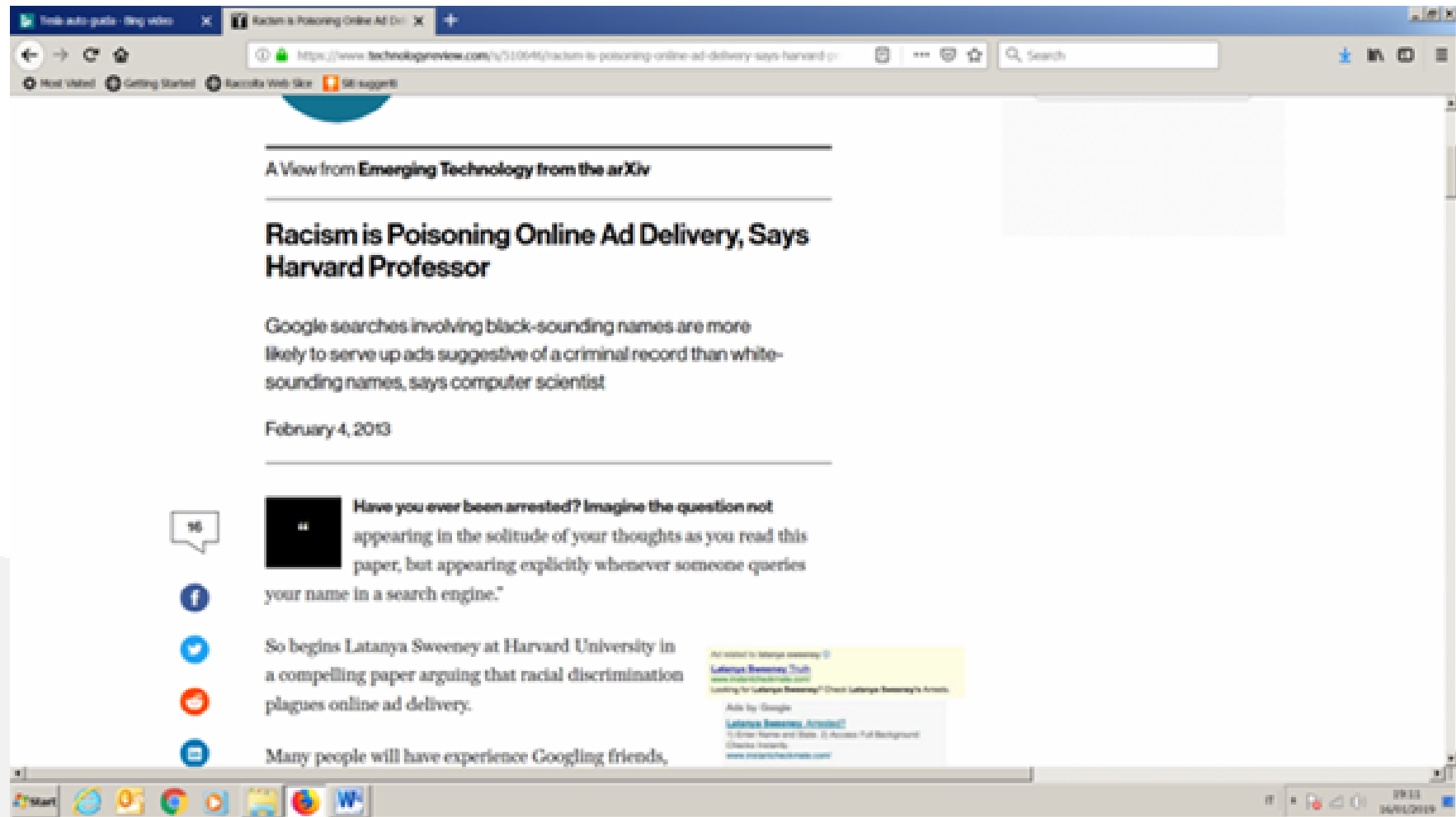
« ... il social network più grande del mondo [Facebook, n.d.r.] aveva messo tra le fasce di clienti che è possibile scegliere anche “**gente che odia gli ebrei**”



[...] ha risposto che **non si tratta di una possibilità** [quella di scegliere gruppi omogenei verso i quali erogare pubblicità, n.d.r.] **creata in modo intenzionale**, ma che **è colpa di un algoritmo**, che ha individuato la categoria da solo e l'ha inserita tra quelle che possono essere target di pubblicità »

L'IA non è esente da colossali cantonate

Google razzista



L'IA non è esente da colossali cantonate

Microsoft ed il suo chatbot razzista

The screenshot shows a web browser window with two tabs. The active tab is titled "Here Are the Microsoft Twitter Bot's Craziest Racist Ra-1766820160" and displays a tweet from Sophie Kleeman. The tweet, posted on 3/24/16 at 11:43am, is titled "Here Are the Microsoft Twitter Bot's Craziest Racist Rants" and features a distorted image of the chatbot Tay. The tweet has 394.3K retweets, 69 replies, and 9 bookmarks. The browser's address bar shows the URL "https://gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160". The left sidebar of the browser shows a "You may also like" section with an article titled "The Dirty Truth About Turning Seawater Into Drinking Water" by Yessenia Funes. The Windows taskbar at the bottom shows the Start button and several application icons, including Internet Explorer, Google Chrome, and Microsoft Word. The system clock in the bottom right corner indicates the time is 19:14 on 16/01/2019.

Microsoft

Here Are the Microsoft Twitter Bot's Craziest Racist Rants

Sophie Kleeman
3/24/16 11:43am • Filed to: GARBAGE FIRES

394.3K 69 9

Facebook

Yesterday, Microsoft unleashed Tay, the teen-talking AI chatbot built to mimic and converse with users in real time. Because the world is a terrible place full of shitty people, many of those users took advantage of Tay's

Share Tweet

Gli errori nell'IA sono inevitabili

Immaginiamo di costruire un algoritmo/macchina (complicati a piacere) che pretendano di simulare alla perfezione i meccanismi del pensiero umano.

Comunque venga realizzata, tale macchina si baserà su un **sistema formale**.

Un simile **sistema formale**, a causa dei **teoremi di Gödel**, si troverebbe nella situazione (prima o poi) in cui **non è in grado di stabilire** se una proposizione sia vera oppure falsa (indecidibilità ed incompletezza).

Uber investe un ciclista

<https://www.youtube.com/watch?v=ufNNuafuU7M>



I limiti dell'IA impongono il problema etico

D.: “Una mappa di dati può essere una copia esatta della realtà?”

R.: “No”.

Le IA lavorano sui dati, ovvero su una descrizione elementare (codificata) di un'entità.

In quanto codifica, il dato è una riduzione della realtà (analoga a quello che fa una mappa geografica, che è una rappresentazione semplificata dello spazio).

Dunque: l'IA prende decisioni su rappresentazioni della realtà.

Questione etica: accettiamo la semplificazione operata dall'IA?

Logica degli approcci etici per l'IA

Approccio UE



OLL19283

S.L.C.

116TH CONGRESS
1ST SESSION

S. _____

To direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments.

IN THE SENATE OF THE UNITED STATES

Mr. WYDEN (for himself and Mr. BOOKER) introduced the following bill; which was read twice and referred to the Committee on

A BILL

To direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments.

1 *Be it enacted by the Senate and House of Representa-*
2 *tives of the United States of America in Congress assembled,*

3 **SECTION 1. SHORT TITLE.**

4 This Act may be cited as the "Algorithmic Account-
5 ability Act of 2019".

6 **SEC. 2. DEFINITIONS.**

7 In this Act:

Approccio U.S.A.

sogei

Sede Legale Via M. Carucci n. 99 - 00143 Roma

Come realizzare un test etico per l'IA

Rispettare il principio del *primum non nocere*

(*principle of non maleficence*).

La realizzazione di tecnologie controllate da sistemi di IA porta con sé una serie di problemi legati al *governo dell'autonomia gestionale*. Gli artefatti IA hanno la capacità di *mutare il loro comportamento* in base alle condizioni in cui operano.

La libertà decisionale deve essere governata da una *artificial moral agent*, ovvero una *moralità artificiale* nei sistemi di IA.

La regola di base degli artefatti IA rispondenti a tale operatività è quella di comportarsi in maniera da *scegliere sempre il male minore* in accordo con l'esperimento ideale circa la situazione del carrello

(cfr. Philippa Foot & Judith Jarvis).

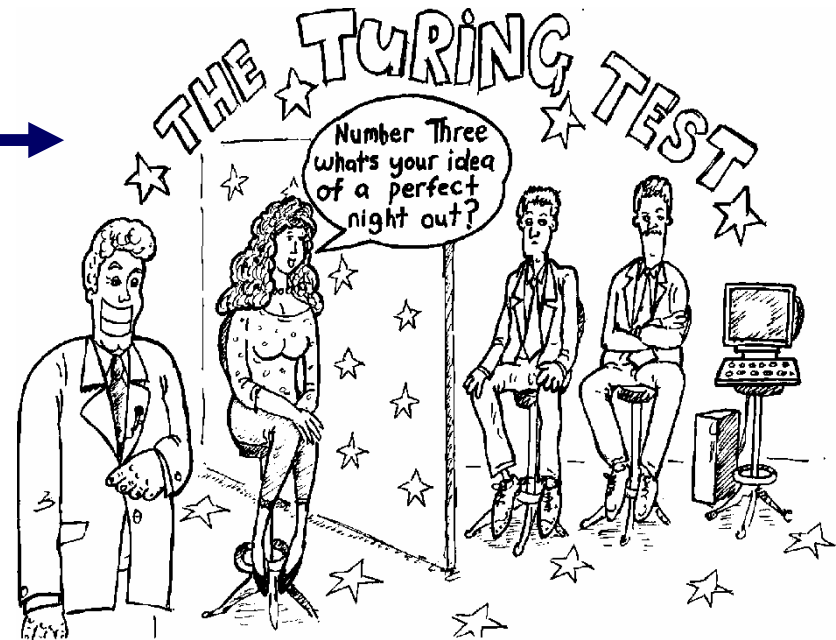


Una proposta per verificare l'etica dell'IA

Invertire il problema:

non regole dal quale far discendere una IA etica, **ma test** per verificare se un artefatto IA corrisponda o meno ai canoni di eticità attesa.

dal “test di Turing” →



← al “test etico”



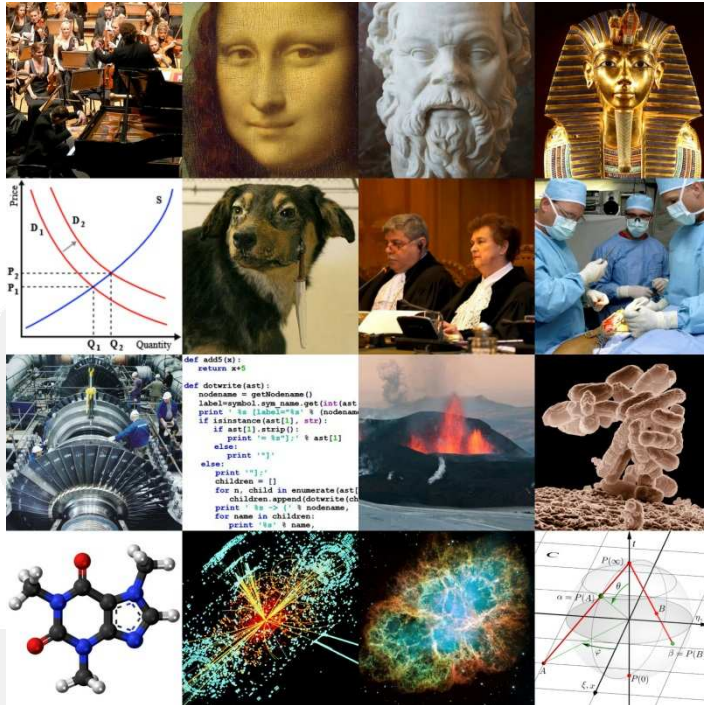
Is it Truthful?
...Authentic?
...Respectful?
...Equitable?
...Socially Responsible?

Then yes, it's ethical!

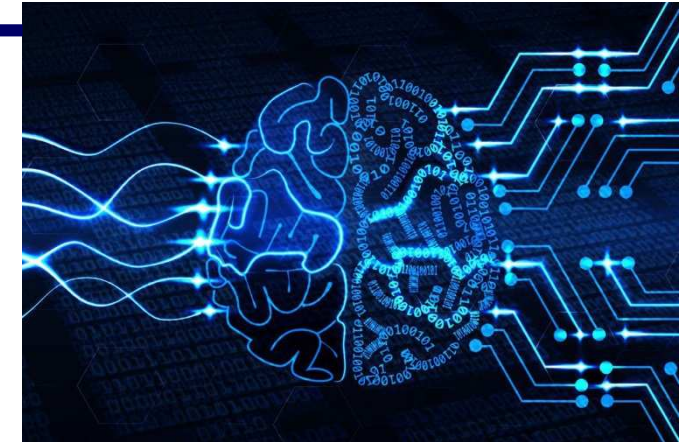
Come realizzare un test etico per l'IA

IA come estensione delle capacità cognitive

(es. credit scoring, giornalista automatico, ecc.)



Regole del test etico definite dalle stesse norme e consuetudini che governano l'ambito cognitivo sul quale agisce l'artefatto di IA



capacità cognitive regolate da norme e consuetudini

